
nature
TRAUMATIC BRAIN INJURY
Consciousness raising therapy
VERTEBRATE ORIGINS
Gene fishing
EATING IN THE GREENHOUSE
Are high-CO_2 crops bad for you?
THE HEAT IS ON
Atmospheric brown clouds enhance climate warming
NATUREJOBS
Atmospheric science

nature
www.nature.com/nature
Vol 448 | Issue no. 7153 | 2 August 2007

Free at last

The liberation of six foreign health workers, held without cause in Libya, is to be welcomed. Now Libya should face facts — and clear their names.

The six medical workers held for eight years in a Libyan prison on false charges of deliberately infecting hundreds of children with HIV were finally freed last week. But Libya's cynical insistence on their guilt is casting a pall over this long-awaited event.

Late in the negotiations that saw the medics' sentences commuted from the death penalty to life imprisonment followed by their extradition to Bulgaria, Libya refused a request for the final settlement to state that it did not represent an admission of guilt. When Bulgaria freed the six, Baghdadi Mahmudi, Libya's prime minister, denounced the pardon as a "betrayal", arguing that the medics should have served life sentences. It is time for Libya to end this charade.

The six's only crime was to be in the wrong place at the wrong time. From the outset they were pawns in a larger geopolitical game in which human rights and justice played second fiddle to *Realpolitik* — in this case, Libya's position as a major oil exporter and its utility as an ally in the 'war against terror'. That the medics are out at all is a tribute to the patience and determination of a handful of European diplomats.

An important supporting role was played by scientists who took up the medics' cause, including Nobel laureate Rich Roberts of New England Biolabs; Vittorio Colizzi, an AIDS researcher at Tor Vergata University in Rome; and Luc Montagnier, whose group in Paris discovered HIV. They all persistently dissected the emptiness of the prosecution case, showed multiple avenues of evidence pointing to a hospital infection as the true cause of the outbreak and campaigned tirelessly.

The scientists quickly learned that effectiveness in such matters demanded tight liaison with defence lawyers and human-rights groups. One-off appeals and letters of protest can have some impact in raising public awareness, but effective advocacy requires sustained action, clear objectives and a strategy to achieve them.

When scientists upped the pressure in the run-up to the trial last autumn (see *Nature* **444,** 146; 2006), calling for the scientific evidence to be heard, some observers argued the approach was naive. After all, the court had consistently refused international expertise in the case. But the strategy had already been recognized by the medics' lawyers and human-rights groups as the best card to play.

Had Libya allowed the scientific evidence to be heard in court, the prosecution case would have collapsed. As was always more likely, it refused this, thus exposing the trial as a sham and providing a useful lever for public, and hence political, opinion.

At other times it was necessary for the scientists and human-rights activists to protect the prisoners' interests by showing discretion in their public statements. It was known from diplomatic sources, for example, that Libya's Supreme Court would uphold the death-penalty verdicts — as it did on 11 July — but that these would be commuted soon afterwards. Had there been huge public outrage at the initial verdict, the commuting of the sentences might have been derailed, so those involved agreed that public reaction should be restrained until the final decision.

> "The six medics were not given a fair trial and scientific evidence that would have exonerated them was ignored."

Now that the medics are free, such restraint is unnecessary. The 1998 outbreak was a triple tragedy: for the six, for the infected children and for human rights. The six were not given a fair trial, prosecution evidence was fabricated and scientific evidence that would have exonerated the medics was ignored. Their trials were a mockery of justice.

Progressive elements within Libya want this truth to come out. Seif al-Islam Gaddafi, son of Libyan leader Muammar al-Gaddafi, played a significant role in resolving the case through one of his charities. He is convinced that the outbreak was an accident, and wants Libya to face up to its AIDS problem and to promote health care.

Libya has, unfortunately, won plaudits in parts of the Arab world for the way it has played its hand, winning normalization of its political and economic ties with the European Union (EU) and much else besides for releasing the six. The EU and the United States should make further normalization contingent on the Libyan government owning up to the real facts of the case, and exonerating the six. ■

Board games

The way research on human subjects is overseen in the United States requires reform.

There is no greater burden of responsibility for scientists than that placed on those who conduct medical research on human subjects. On the rare occasions that this duty is inappropriately discharged, the results can be devastating. Even so, once the initial outcry dies down, little tends to change.

The diverse collection of institutional review boards (IRBs) that oversee such research in the United States barely qualifies as a 'system'. Despite repeated attempts by the Institute of Medicine and others to highlight their shortfalls, the quality and effectiveness of the boards remain patchy (see page 530).

As committees struggle with heavy caseloads, their ability to monitor ongoing trials is weakened. A large research hospital can process hundreds of applications per year, and gets little help from the federal government. The Office for Human Research Protections oversees thousands of local ethics committees and billions of dollars' worth of clinical research, and operates on an annual budget of just $7 million.

If the US government wanted to strengthen the way human clinical trials are overseen, adequate funding for the Office for Human

Research Protections would be an obvious place to begin. Another improvement would be more widespread accreditation of the IRBs, to help ensure proper training and support for committee members.

But further reforms are also necessary. Some see centralizing the review of multicentre studies as a way of relieving the burden on local review boards. But the debate over whether to centralize, or even regionalize, the review of studies is complex. The United States is a large, diverse and, most of all, litigious country, and local boards help universities to fend off legal action by showing that they have taken responsibility for what goes on within their walls. Nonetheless, centralization deserves to be explored further.

One approach would set up national committees, perhaps run by the National Institutes of Health, that could establish clearer guidelines on the ethical quandaries commonly faced by local review boards. Questions over payments to volunteers, for example, or on what constitutes informed consent, need not all be answered on the current ad hoc basis.

Another challenge facing US authorities is the fact that not all research on human subjects is overseen by the federal government. The IRBs are not federal, but federally funded research on human subjects must be reviewed by them, as must any trial that becomes part of a submission to the Food and Drug Administration. Some states and institutions also require all research on human subjects to be vetted by an IRB. That leaves room for privately funded research to proceed without any requirement for ethical review. It isn't known how much research on human subjects occurs without review by government officials or IRBs. But such research may expose patients to unnecessary risks.

In 1999, the review-board system came under some scrutiny when 18-year-old Jesse Gelsinger died during a gene-therapy trial at the University of Pennsylvania in Philadelphia. In 2001, a healthy woman died after taking an unapproved asthma medication during a clinical trial at Johns Hopkins University. And looking farther afield, in London last year six men became seriously ill during a clinical trial of a monoclonal-antibody therapy.

Overall, it is not a bad record. But that is as much by luck as by design, and if it is to remain the case, real improvements need to be made to the IRB system. The impetus for such change should not have to rely on the bursts of interest that tend to follow mishaps during human trials. ■

A sporting chance

Bans on drug enhancement in sport may go the way of earlier prohibitions on women and remuneration.

Whether you have been following the just-finished Tour de France or waiting for Barry Bonds to break the all-time record for major-league home runs in baseball, the topic of drugs in sport has been hard to avoid of late.

To cheat in a sporting event is a loathsome thing. For as long as the rules of the Tour de France or any sporting event ban the use of performance-enhancing drugs, those who break the rules must be punished whenever possible. But this does not preclude the idea that it may, in time, be necessary to readdress the rules themselves.

As more is learned about how our bodies work, more options become available for altering those workings. To date, most of this alteration has sought to restore function to some sort of baseline. But it is also possible to enhance various functions into the supernormal realm, and the options for this are set to grow ever greater.

The fact that such endeavours will carry risks should not be trivialized. But adults should be allowed to take risks, and experience suggests that they will do so when the benefits on offer are enticing enough. By the end of this century the unenhanced body or mind may well be vanishingly rare.

As this change takes place, we will have to re-examine what we expect of athletes. If spectators are seeking to reset their body mass index through pharmacology, or taking pills that enhance their memory, is it really reasonable that athletes should make do with bodies that have not seen such benefits? The more the public comes to live with the mixed and risk-related benefits of enhancement, the more it will appreciate that allowing such changes need not rob sport of its drama, nor athletes of their need for skill, training, character and dedication.

To change the rules on pharmacological enhancement would not be without precedent. It was once thought that a woman could not epitomize the athletic ideal as a man could, and so should be stopped from trying. Similarly, it was thought proper to keep all payments from some athletes, thus privileging the already wealthy. These prejudices have been left behind, and the rules have changed. As pharmacological enhancement becomes everyday, views of bodily enhancement may evolve sufficiently for sporting rules to change on that, too.

> **"Is it really reasonable that athletes should make do with bodies that have not been enhanced?"**

This transition will not be painless. Some people will undoubtedly harm themselves through the use of enhancements, and there would need to be special protection for children. That said, athletes harm themselves in other forms of training, too. They may harm themselves less with drugs when doctors can be openly involved and masking agents dispensed with.

There is also the problem of who goes first. The first sport to change its rules to allow players to use performance-enhancing drugs will be attacked as a freak show or worse. The same may be true of the second. This may well have the effect — may already be having the effect — of delaying the inevitable.

Perhaps the Tour de France could show the way ahead here. In terms of public respect, endurance cycling has the least to lose and perhaps the most to gain. To be sure, a change in the rules would lead to the claim that 'the cheats have won'. But as no one can convincingly claim that cheats are not winning now, or have not been winning in the past, that claim is not quite the showstopper it might seem to be.

A leadership ready to ride out the outrage might be better for the sport in the long run. If some viewers and advertisers were lost along the way, the Tour could console itself with the thought that it got by with far less commercial interest in days gone by — and that it is more likely to re-establish itself through excellence and honesty than in the penumbra of doubt and cynicism that surrounds it now. ■

RESEARCH HIGHLIGHTS

Seismic storms

Geophys. Res. Lett. **34,** L14304 (2007)

Thunder really does make Earth move. The acoustic shockwaves unleashed by a thunderstorm can excite seismic tremors in the ground, say Ting-Li Lin and Charles Langston of the University of Memphis in Tennessee.

It is not the audible crash that does most of the shaking, however, but the associated rumble on the threshold of hearing, in the low-frequency infrasound range. Lin and Langston used a seismic sensor array housed in 0.2-metre-deep boreholes to pick up the tremors created by two thunderclaps in the spring of 2006 . They say that thunder could provide a convenient natural source for mapping out the seismic vibration profile of an earthquake monitoring site.

MEDICAL MODELS

Humanized livers

Nature Biotechnol. doi:10.1038/nbt1326 (2007)

Scientists have constructed an improved mouse model for studying human liver cells.

The liver breaks down foreign substances such as drugs. Studying its cells gives insight into drug toxicity as well as liver diseases such as hepatitis. But it's difficult to culture human liver cells, so scientists have created mouse models whose own liver cells die to 'make space' for transplanted human cells.

Previously available models suffered from practical difficulties, so Markus Grompe of the Oregon Health and Science University in Portland and his colleagues set about making a better one. They bred mice deficient in a key liver enzyme, and then treated them with a protein that further damages their livers. In these mice, transplanted human cells taken from live patients or cadavers multiplied extensively. Cells harvested from these 'humanized livers' could then be cultured or serially transplanted into other mice for further expansion.

NITROGEN CYCLE

Attack of the nitrogen fixers

Global Change Biol. doi:10.1111/j.1365-2486.2007.01410.x (2007)

Scruffy green invaders have colonized the islands of Hawaii and are boosting their production of greenhouse gases, new research confirms.

The invader is a tree named *Morella faya* (pictured right). This plant's ability to convert atmospheric nitrogen into biologically useful compounds has given it a competitive edge over many native species. Previous work has suggested that *M. faya* has been increasing local soil nitrogen cycling and the concentration of nitrogen in the forest canopy.

Sharon Hall of Arizona State University in Tempe and Gregory Asner of the Carnegie Institution in Stanford, California, expanded on this work by coupling measurements of soil nitrogen with remote sensing of upper-canopy nitrogen oxide concentrations in Hawai'i Volcanoes National Park. These concentrations imply that the tree has pushed up regional nitrous and nitric oxide emissions 16-fold since its first occurrence in the park 40 years ago.

CHEMISTRY

Getting the right version

Science **317,** 496–499 (2007)

Mirror-image versions of some molecules — called chiral — are not structurally identical, and ensuring that only one version is produced in a chemical reaction is often a problem. One way is to use a metal catalyst attached to a ligand molecule that is also chiral.

Ligands are normally attached to metal catalysts by strong covalent bonds, but Dean Toste and his colleagues at the University of California at Berkeley have successfully used negatively charged ligands that are only weakly ionically attracted to the positively charged metal. This means that a small group of charged ligands could make a vast number of existing metal-catalysed reactions chirally selective.

Toste reports three different gold-catalysed transformations that gave more than 90% yield of one chiral product.

BIOMEDICINE

Be still my beating heart

Cell **130,** 247–258 (2007)

Heart-pounding action may pack a thrill, but researchers have shown that getting rid of a protein involved in adrenaline's control of heart rate allows mice to live longer and have healthier hearts. This protein, known as type 5 adenylyl cyclase (AC5), could become a new target for anti-ageing therapies.

Together with their colleagues, Stephen Vatner and Junichi Sadoshima at the New

Jersey Medical School in Newark found that mutant mice lacking AC5 lived, on average, 30% longer, weighed less, and exhibited less age-induced heart stress and bone loss. Mutant mice also produced more of the antioxidant enzyme superoxide dismutase.

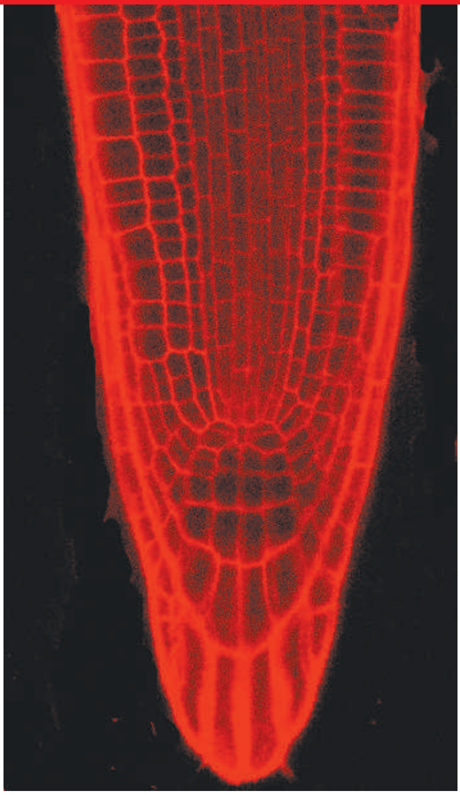
BOTANY

At the root of it all

Science **312**, 507–510 (2007)

In plants, ethylene responds to environmental signals such as being touched. Now, ethylene has also been discovered to relay these signals to the very heart of the root's stem-cell production factory.

Inside the root tip (pictured right), a group of four cells known as the quiescent centre (QC) gives rise to all the cells in the tissue systems of the root. Liam Dolan at the John Innes Centre in Norwich, UK, and his colleagues, compared wild-type *Arabidopsis thaliana* (thale cress) with mutants that had defective synthesis of ethylene. They found that ethylene promotes cell division in the QC to create further stem cells. But it does not induce them to differentiate. The latter task is more closely associated with the hormone auxin, and the two chemicals work in concert to orchestrate root growth.



ASTROCHEMISTRY

Anions in space

Astrophys. J. **664,** L43–L46 (2007);
Astrophys. J. **664,** L47–L50 (2007)

Astronomers using a 100-metre radio telescope have found the largest negatively charged molecule yet in space. According to the researchers, the discovery of octatetraynyl anion (C_8H^-) and three other anions in the past year offers intriguing evidence for a suite of chemical reactions and products not yet observed — including molecules similar to amino acids and other precursors of life.

A team at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, spotted the compound in the dark molecular cloud TMC-1 in the constellation Taurus. And researchers at the National Radio Astronomy Observatory in Charlottesville, Virginia, have found it in the halo of IRC+10 216, a dying star in Leo.

SEX DETERMINATION

Old sperm sires sons

Biol. Lett. doi:10.1098/rsbl.2007.0196 (2007).

Older sperm are more likely than fresh sperm to produce males, at least in lizards. The discovery comes from a species in which females store sperm after copulation for later use, and it may explain why this species' sex ratio skews towards males as the reproductive season progresses. Sperm stored in the females' reproductive tract often outcompetes sperm from more recent inseminations.

A team of zoologists, led by Mo Healey of the University of Wollongong, Australia, set up breeding pairs of a small lizard called the Australian painted dragon (*Ctenophorus pictus*) in the laboratory. Stored sperm produced 55% sons, whereas fresh sperm sired only 32%.

NANOTECHNOLOGY

The look and feel of nano

Nature Nanotechnol. **2,** 407 (2007)

Ozgur Sahin of Harvard University and his colleagues have developed a specialized atomic-force microscope (AFM) that 'feels' a surface's softness. Existing AFMs create images by passing a vibrating cantilever with a sharp tip across a material. Sahin's team used a T-shaped cantilever that twisted as it vibrated up and down. The twisting changed in response to the material's softness, providing more detailed information than conventional tools allow.

Sahin says the technique shows promise for a variety of nanotechnology and biotechnology procedures, such as probing the mechanical properties of proteins and DNA.

Correction

In the Research Highlight 'Assorted fungi' (*Nature* **447,** 1034; 2007) there was an error in the reference. The correct details are: *Mycol. Res.* **111,** 509–547 (2007).

JOURNAL CLUB

Paul Kenrick
The Natural History Museum, London, UK

A palaeobotanist finds answers to the origin of roots in the genes of a living moss.

Roots have been called the hidden half of plant diversity. Confined mainly to the subterranean, their unseen influence extends well beyond the plant that they sustain to form an integral component of soil ecosystems and a significant link in the carbon cycle.

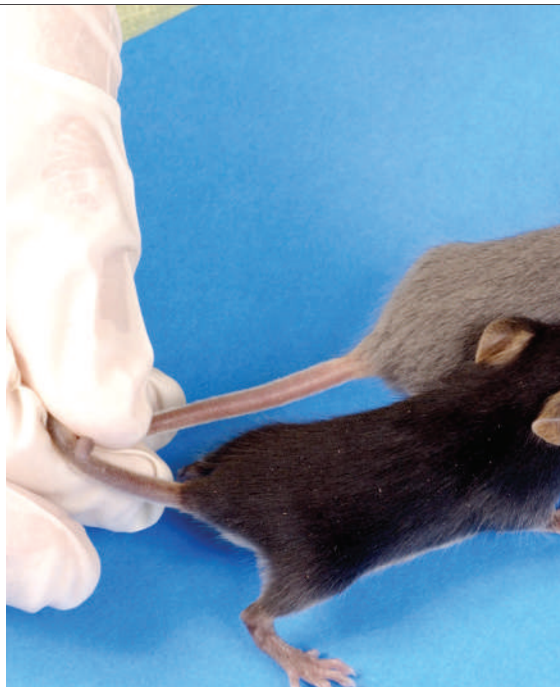
In my research, I use fossils to piece together how the fundamental organs and basic lifecycles of plants evolved, and roots are one of the key systems. The fossil record shows that roots were an early innovation in the colonization of the land, and that they evolved remarkably rapidly, developing a diversity of forms comparable to those of the aerial shoots, stems and leaves. Comparative morphology is good for documenting how roots evolved, but are there any underlying molecular developmental similarities among the rooting structures of early plants?

An elegant piece of recent research shows that a similar transcription factor encoded by the gene *ROOT HAIR DEFECTIVE 6* regulates root-hair development in the flowering plant *Arabidopsis thaliana* and rhizoid development in the moss *Physcomitrella patens* (B. Menand *et al. Science* **316,** 1477–1480; 2007). Because flowering plants and mosses diverged more than 400 million years ago, this surprising result implies that the cells with a key role in nutrient acquisition and anchorage in most land plants share a molecular developmental pathway that is very ancient indeed.

More surprising still is the notion that these genes are expressed in both haploid and diploid plants — that is, those whose cells have one or two sets of chromosomes, respectively. Many plants cycle between haploid and diploid forms during their lifecycles. Menand *et al.* propose that genes expressed in early haploid plants were turned on in many tissues during the evolution of plants with diploid phases. Pending further testing, this interesting model is plausible for components of the vascular system, cortex, epidermis, shoot and root.

NEWS

The mouse map gets a lot more signposts



Signs of diversity: coat colour was the first way in which different mouse strains were recognized.

That researchers' staple, the laboratory mouse, has undergone the ultimate dissection: two competing teams have completed the most thorough maps yet of its genetic variation.

The catalogue of 8.27 million genetic differences provides a unique insight into the evolutionary origins of the mouse, and researchers also anticipate that it will speed up the discovery of disease-causing mutations in mice — and ultimately in humans.

Although the two groups disagree over the evolutionary history of the lab mouse (*Mus musculus*), they both assert that the maps will give researchers fast access to an unparallelled cache of genetic information on hundreds of lab strains. "It's a whole new era of mouse genetics," says Gary Churchill, a mouse geneticist at Jackson Laboratory in Bar Harbor, Maine, and an author on one of the studies (H. Yang, T. A. Bell, G. A. Churchill & P.-M. de Villena *Nature Genet.* doi:10.1038/ng2087; 2007).

"It's a whole new era of mouse genetics."

The researchers traced the lineage of small chunks of the mouse genome in 11 common lab strains back to four wild subspecies. Because of the genetic similarity between the 450-odd inbred classical lab strains, a relatively small number of single-nucleotide polymorphisms (SNPs), which are one-base-pair differences between two strains, can now be used to deduce the sequences of more than 8 million SNPs using the new map. If one segment of the genome can be attributed to a single ancestor, then all mouse strains that share that segment should have the same SNPs. As few as 50,000 SNPs spread across the genome per strain are needed to deduce the rest of the 8 million.

"You don't have to go and resequence [their genomes]," says Richard Mott, a bioinformatician at the Wellcome Trust Centre for Human Genetics in Oxford, UK, who was not involved in the studies. "That's tremendous."

Mus musculus is the most important model organism for human disease and the standard-bearer in cancer and immunology research, as 99% of mouse genes have a human counterpart. In the early 1900s, Abbie Lathrop, a former schoolteacher in Massachusetts, started breeding mice obtained from fanciers in Europe and Asia. These animals spawned the first lab, which were made genetically pure through dozens of generations of sibling mating. "It was really, truly just a handful of mice," says Kelly Frazer, a geneticist who led the second mapping effort while at the biotechnology company Perlegen Sciences in Mountain View, California (K. A. Frazer *et al. Nature,* doi:10.1038/nature06067; 2007).

Subtle differences between strains — A/J mice are prone to cancer and BALB/cJ mice are susceptible to the bacterium *Listeria*, for example — can be traced to SNPs, of which fewer than 45 million are thought to exist in lab and wild strains together.

Difference matters

The first mouse genome to be sequenced was that of C57BL/6J, in 2002. The US National Institute of Environmental Health Sciences in Triangle Park, North Carolina, teamed up with Perlegen to roughly sequence 10 more strains, plus four subspecies of wild mice collected from all over the world. The sequences were completed in 2006 and made freely available.

To tease out more detail and to make the information applicable to other mouse strains, Frazer's team made a genomic map of each sequenced strain, and established the ancestral subspecies that gave rise to around 40,000 segments of the genome. Overall, says Frazer, a subspecies from Europe called *domesticus* contributes most (68%) of the lab mouse genome, but this varies from chromosome to chromosome and strain to strain.

Frazer's team didn't know that Churchill's team was doing the same thing. "It would have been nice if the two groups had talked with one another," she says, "but they didn't let us know they were doing it."

Using a slightly different method, Churchill and his colleagues at the University of North Carolina at Chapel Hill also conclude that *domesticus* provides the lion's share of diversity in lab mice, more

Mice unlimited

Imagine a laboratory that is home to a thousand strains of mice, maintained by a dedicated staff and visited by scientists seeking access to the most diverse collection of lab mice in the world. This is the vision of geneticist Gary Churchill at Jackson Laboratory in Bar Harbor, Maine. "It's a new model, and I don't think biologists have embraced it yet," he says. Yet a loosely knit team of hundreds of researchers all over the world called the Complex Trait Consortium is already at work providing the mouse strains that would fill such a facility.

Launched in 2005, the consortium's Collaborative Cross project seeks to quadruple the genetic diversity represented by the 450 or so classical mouse lab strains. More mice with relevant phenotypes will help geneticists winkle out the genes and mutations that underlie complex traits and diseases.

Currently, huge swathes of the genome differ little between lab strains — genetic blind spots that complicate attempts to associate genes with disease. Reshuffling the mouse gene deck will eliminate such blind spots, says consortium member David Threadgill, a geneticist at the University of North Carolina in Chapel Hill.

Hosted by Oak Ridge National Laboratory in Tennessee, the breeding project is just getting under way with eight strains of mice being crossed in hundreds of combinations.

Each combination will be interbred for at least 25 generations to ensure new strains are pure-breeding. Already, 400 lines have been bred for 10 generations and those strains should be ready within three years, says Churchill. **E.C.**

MEDICAL OPINION COMES FULL CIRCLE ON CANNABIS DANGERS
Frequent use more than doubles psychosis risk.
www.nature.com/news

JACKSON LABORATORY

than 90%. He says that the mouse genome is far more complex than he had thought. "This has been an unfolding story," he says. "We saw patterns that were so confusing." What has emerged is only a partial sketch, he says. Sequencing the entire genomes of more laboratory and wild strains will paint a fuller picture and settle the differences between the two maps. "It's going to be resolved," Churchill says.

Although the teams present two evolutionary histories for the lab mouse, "the similarities are more important than the differences," Mott says. And the ability to deduce polymorphisms in other strains doesn't depend on an unambiguous history.

Researchers will be eager to make use of the data, say both teams. Mark Daly at the Broad Institute in Cambridge, Massachusetts, and a co-author with Frazer, has already applied the map to 100 strains of mice that had been genotyped for 150,000 SNPs, and will soon be able to assign millions more SNPs to these mice.

The projects could also spur the use of less common lab strains by providing better genetic maps, says Elizabeth Fisher, a geneticist at University College London. "There's an enormous amount of diversity out there that we're not capitalizing on," she says. A collaboration of hundreds of researchers called the Complex Trait Consortium plans to breed 1,000 new mouse strains (see 'Mice unlimited'). And now those strains will be easily genotyped. "It's going to allow us to make better mouse lines for the future, with levels of diversity that are more like human diversity," Churchill says. ■

Ewen Callaway

Q&A: Jürgen Graeser

On 29 August, Jürgen Graeser of the Alfred Wegener Institute for Polar and Marine Research in Potsdam, Germany, will embark on an eight-month Russian-led voyage on a d rifting ice floe to the North Pole. The German will join 35 Russian colleagues from the Arctic and Antarctic Research Institute in St Petersburg, Russia, to collect a wealth of atmospheric and sea-ice data in one of the most remote and inhospitable regions of the planet. **Quirin Schiermeier** spoke to the first Westerner invited to spend time on a Russian drifting station.

Why is this expedition special?
There'll be no ship, no aircraft, no fixed route, nothing. We will never know what the next day will bring. For me, this is the ultimate challenge. Work in Arctic stations has, of late, become quite comfortable. I've read many books about what 'real' expeditions were like, but I never thought it would be possible to participate in such an adventure.

Where will you be at the winter solstice?
I'll be drifting in the Arctic Ocean, somewhere near the North Pole, on an ice floe some 2 kilometres across and 3 metres thick. A suitable floe is being sought now. The Russians will build a makeshift research base on it made of prefabricated wooden cabins.

What will your daily routine be?
I'll be in charge of a small 'aerological' station, doing a range of atmospheric and meteorological measurements. I'll also use radiosondes and a tethered balloon to make a profile of the ozone and aerosol concentrations in the lowest layer of the atmosphere. Such high-resolution measurements have never been taken before at such high latitudes. At the end of each day, I'll transmit the most basic data to Germany. Satellite communication beyond 80° N is extremely slow, so this will take some time. Other than that, the days will be punctuated by the meals we'll eat together.

Russia is claiming parts of the Arctic. Is there a political background to the project?
The expedition is not equipped to search for geological evidence that could back up such claims. During preparatory meetings, the plans for exactly what the Russian team will do have had a low profile, but I doubt that any ill will is intended. The Russians have been using drifting stations for more than 50 years, and this is the first time a foreigner has been invited, which signals improved scientific openness towards the West. There's an agreement that all data will be made openly available as well.

What can go wrong?
You need to be constantly aware of polar bears. All team members have been trained to shoot. Also, the floe might become unstable or break. The worst case would be if the floe were to sink. This happened in 2004, and the whole team had to be evacuated by helicopter. Alcohol is also a bit of a concern: each member of the team is officially permitted one can of beer and one bottle of spirits per week. I would have preferred a bit more beer and less spirits. Drink can have disastrous consequences there. ■

Jürgen Graeser with a tethered meteorology balloon.

H. GERICKE/ALFRED WEGENER INSTITUTE

SPECIAL REPORT

China's green accounting system on shaky ground

Beijing is attempting to introduce a 'green GDP' accounting scheme as a way of ranking local governments' success in achieving economic growth that does not cost the environment. But the scheme is faltering at the pilot stage due to political infighting, says **Jane Qiu**.

Beijing has stifled an official report that measures the economic cost during 2005 of China's pollution and environmental damage — its 'green' gross domestic product (GDP) — claiming the methods and data available to make such a calculation are too crude for the findings to be meaningful.

Critics fear that this may have sent out a dangerous signal for local authorities to continue pursuing economic growth at all costs.

Originally scheduled for release in March, the "indefinite postponement" announced last week of the second report of the pilot scheme follows fierce dispute between China's environmental watchdog, the State Environmental Protection Administration (SEPA) and the National Bureau of Statistics of China (NBS).

At a press conference on 12 July, Xie Fuzhan, director of NBS, openly attacked the scheme, questioning its feasibility and methodology. Because of these uncertainties, it should be submitted to the State Council (the country's chief administrative authority) for reference, rather than released to the public, he told journalists. Wang Jinnan of the Chinese Academy of Environmental Planning (part of SEPA), who leads the team calculating green GDP, held a press conference at which he defended the project and announced that the report would not be released anytime soon owing to disagreement between the two administrations. He also blamed the report's suppression on resistance from some local authorities, which he said had been lobbying the central government against it in their eagerness to maintain economic growth.

It is no secret that SEPA and NBS have been arguing over the new accounting scheme since its launch in 2004, but many fear that their increasingly public turf war undermines SEPA and may lead to the accounting system being abandoned altogether.

"A conservative estimate of China's environmental problems."

Although China has stated its commitment to environmental protection since 1983, passing more than 100 environmental laws, economic development traditionally takes priority and local officials in China are still judged largely on their ability to improve economic growth. The pilot scheme — which subtracts the cost of natural resources used and pollution caused from regular GDP — aimed to mitigate some of these excesses and help guide new development.

In September 2006, the central government released its first green GDP report with much fanfare. Internationally, it was greeted as a significant step forward by China in tackling its environmental problems. The report found that pollution cost China 511.8 billion yuan (US$62 billion at 2004's official exchange rate, which underestimates the true economic value of the yuan), accounting for 3% of GDP for 2004. The environmental cost of water pollution, air pollution, and solid waste and pollution accidents accounted for 55.9%, 42.9% and 1.2% of the total costs, respectively.

The scheme does have critics — for example, it does not account for groundwater or soil contamination. "On the whole, this accounting system gives a conservative estimation of China's environmental problems," says Lei Ming, an environmental economist at the Guanghua School of Management, Peking University, who helped compile the report.

Admitting that the method of calculating green GDP needs to be improved, Lei says: "One must take a realistic and pragmatic approach towards the system. We can't get everything right at once and things will improve over time." He thinks that publicising the data could help curb environmental damage. "The stakes are high and we don't have a better alternative at the moment," he says.

Yang Xin, president of Green River, a Chinese non-governmental organization, agrees. He is "bitterly disappointed" at the report's suppression, saying: "An unprecedented environmental crisis is unfolding in China. We can't

Monitoring air pollution

The State Environmental Protection Administration (SEPA) has earmarked 4 billion yuan over the next two years to develop air-monitoring systems for the four main pollutants — sulphur dioxide, nitrogen dioxide, particulate matter measuring 10 microns or less and carbon monoxide. By the end of 2008, all provinces must establish air-monitoring centres and the provinces surrounding Beijing must have the systems running by the end of 2007 in preparation for the 2008 Olympics.

"It's a huge amount of money. If after a couple of years air quality is not much better we cannot say this is not our responsibility," says Liu Bingjiang, head of SEPA's air pollution division. Given China's size, it will be a challenge.

According to Jiming Hao, an environmental scientist at Tsinghua University in Beijing, only 47 of 600 major cities in China currently have adequate air-monitoring systems. There is talk of adding ozone to the list of pollutants to be tested for, but so far Beijing only has seven ozone monitoring stations, compared with 30 for monitoring the other four.

SEPA will train personnel to assess the anti-pollution technologies because factories have installed equipment to meet regulations, but don't always use them owing to running costs.

David Cyranoski

2007 has been named China's year of algae because of repeated outbreaks due to pollution.

SINGLE GENE DELETION BOOSTS LIFESPAN
Mutant mice live longer, age slower and eat more.
www.nature.com/news

afford to wait for a perfect accounting system to emerge while watching local governments destroying our environment."

However, some environmental groups are sceptical of the scheme. "It is premature to implement such a system due to lack of an established formula to follow," says Zhang Jianyu, programme manager at the Beijing office of the Environmental Defense, a US non-profit environmental advocacy group. "There are other ways of measuring environmental impact."

Technical hurdles

Undoubtedly, there are technical hurdles in setting up a method of calculating green GDP. The United Nations, which has been working on a similar system for integrating environmental and economic accounting since the early 1990s, is still debating with economists, environmentalists and policy makers on how it should be done. According to Karl Schorer, head of the environmental-economic accounting department at the Federal Statistical Office of Germany, the main obstacle of such systems is the difficulty in calculating the market value of environmental impact — be it the extinction of a species, the cost of soil erosion due to deforestation, or the health damage from pollution.

In some developed countries such as Germany, the current practice is to associate specific environmental damage — for example, carbon dioxide emissions — with the economic activities that cause it, and to have polices in place to regulate the industries involved. Alternatively, one can set a environmental target, for example, cutting emissions, and then calculate the estimated cost to reach that target. "It is very challenging to take everything into consideration and come out with a single figure," says Schorer.

However, some critics claim that scientific difficulties are beside the point, the NBS is succumbing to pressure from local authorities. "The methodological issue is just a pretext," says a Beijing-based environment writer who asked not to be named. Sources close to SEPA told *Nature* that the scheme has met with fierce resistance from some local governments and that these are largely responsible for the difficulties it is facing.

"Some officials turn pale at the mention of green."

"Green GDP is a highly sensitive topic at the local level," says Lei. Some officials "turn pale at the mention of green." The report's suspension "shows that the scheme has touched a nerve," he adds.

Jennifer Turner, coordinator of the Washington-based China Environment Forum, which acts as a clearing house for information on energy and environment issues in China, says the refusal by local governments to participate in the green GDP programme shows they are afraid of more information escaping. "The squirming of the local governments means [the green GDP plan] is working. They're running scared. Their feet are being put to the fire."

The second report ranks all 31 provinces, autonomic regions and municipalities according to their green GDP. Some areas, especially those in the central and western part of the country, are relatively high on the GDP list, but their rankings fall significantly when environmental costs are factored in, says a SEPA source. An article published in the *Economic Observer*, a Chinese newspaper, claims that the second report shows that there was an increase in the proportion of economic loss due to environmental pollution as relative to GDP in 2005 compared with the previous year.

Lei declined to confirm the report, but says that "you don't need the numbers to tell that the environmental problems in China are worsening". Indeed, they are painfully visible: Beijing, which is set to host the 2008 Olympics, has reported its worst June air pollution in seven years. Meanwhile, this year has seen repeated outbreaks of blue-green algae in the country's lakes and waterways — so much so that 2007 has been named China's year of algae. According to an article in the *Financial Times* this month, China asked the World Bank not to publish "sensitive" estimates of the number of premature Chinese deaths each year from polluted air and water, calculated to be around 460,000, with a further 300,000 who died from indoor toxins.

The 'green' price of development can be scarily high. In the northern province of Shanxi, a coal production centre, environmental costs and natural-resource losses constituted a third of its GDP in 2002. "Of course, local officials don't want to see that on paper," says the SEPA source.

The timing is a crucial factor, too. The Chinese Communist Party will hold its 17th five-yearly party congress this autumn, which will involve a lot of reshuffling, and officials would not like to have green GDP pose a stumbling block to their promotion.

A final decision on the fate of the green GDP accounting system will be made by the State Council, and all parties involved are holding their breath.

As a small ministerial authority, SEPA has long suffered from lack of resource and actual power. Although authorized to regulate the country's environment, it's often an uphill struggle for the agency to exert its influence among local governments and the business of entities on a par with itself. If the green GDP scheme were brought to a close, SEPA would have a tougher time ahead. ■

Jane Qiu is a freelance writer based in China.

GETTY

J. YU/REUTERS

SCORECARD

Stem cells on the ISS NASA's decision to include privately funded research on the International Space Station from 2011 could see stem cells, and other biomed products, going into space.

Everything else at NASA With charges of drunken astronauts and lost and sabotaged equipment, NASA has more pressing concerns to address before turning to next decade's bioethics issues.

ON THE RECORD

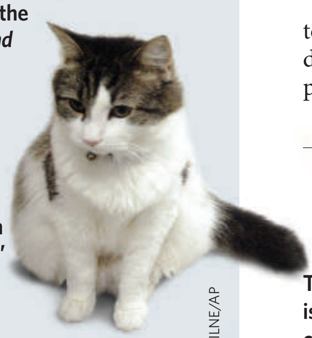
"The toes date from between 1000 and 600 BC, so if we can prove they were functional we will have pushed back prosthetic medicine by as much as 700 years."

Biomedical egyptologist Jacky Finch of the University of Manchester, UK, who is studying ancient Egyptian big toes, made of wood and leather, from museums in Cairo and London. She's looking for people who have lost their right big toe and wouldn't mind helping to test out a modern replica.

OVERHYPED

Oscar, the cat of doom

If you believe the news, the cat who can "predict death" at a Rhode Island nursing home has "left doctors baffled". And so, abetted by the *New England Journal of Medicine,* where the story first appeared, summer's silly season for 'science' stories begins.



S. MILNE/AP

ZOO NEWS

Tentacle treasure

An octopus found with a porcelain plate stuck to its suckers has led archaeologists to a hoard of ancient pottery shipwrecked in a twelfth-century boat off the coast of South Korea.

Sources: The Scientist, news@nature.com, EurekAlert, BBC, N. Engl. J. Med.

SIDELINES

Russia at forefront of Arctic land-grab

Marine scientists are voicing concerns that their research may be jeopardized by the international scrum to claim the Arctic seabed.

Last week, Russia highlighted its claims to a vast chunk by sending a mission to drop a titanium capsule with a Russian flag on the seabed underneath the North Pole. Canada, Norway, the United States and Denmark (on behalf of Greenland) are also preparing claims.

The 1982 United Nations Convention on the Law of the Sea (UNCLOS) allows states an economic zone that extends 200 nautical miles from their coastline. To increase this, countries must prove to the United Nations Commission on the Limits of the Continental Shelf that their physical continental margin extends farther than this. If Moscow can prove that the structure of the continental shelf under the ocean is geologically similar to that of Russian land, it may be able to extend its territory. The Lomonosov Ridge, which it sees as a continental extension, is key to its claims. Russia hopes to claim 1.2 million square kilometres and, with it, the rights to copious mineral and fossil-fuel reserves. Estimates vary, but some suggest that the Arctic contains as much as 25% of the world's oil reserves.

A successful claim would not allow nations to control water above the seabed, but anyone doing research on the ocean floor would need permission, which worries some involved in geological and polar research. "Access is already a problem," says Mead Treadwell, chair of the US Arctic Research Commission. "We've looked at maps which show there could be very little free space left in the Arctic Ocean."

Anders Karlqvist, director of the Swedish Polar Research Secretariat, agrees. "It will probably mean research will be more regulated in terms of getting permission and documentation that goes through various organizations in the countries concerned. And if the political climate changes it may lead to research restrictions."

"There could be very little free space left in the Arctic Ocean."

But the contest is providing plenty of work for some marine geologists in the run-up to many countries' deadline for submissions of May 2009. "I've never worked so hard on strategic policy as now," says Lindsay Parson, who leads the UNCLOS Group at the United Kingdom's National Oceanography Centre in Southampton.

To prove their claims, countries need to collect bathymetric data that accurately describe the underwater lie of the land and seismic data on the structure of the sea floor. These data can be used to determine where the continental shelf off a country's coast falls away in a 'continental slope' and how far sediments derived from the land extend out beyond the foot of that slope. The sea floor can be claimed if it is sedimentary rock at least 1% as thick as its distance from the foot of the continental slope. These measure-

New life for nuclear warheads

The US Department of Energy is starting construction of a controversial facility to recycle nuclear warheads.

On 1 August, work finally began on a US$4.8 billion mixed-oxide fuel-fabrication facility at the Savannah River Site in South Carolina. The plant will convert surplus, weapons-grade plutonium into fuel for sale to commercial power reactors in North and South Carolina.

Supporters claim that the facility will help to dispose of dangerous weapons material safely. But sceptics, including some key members of Congress, balk at its enormous cost and question whether it will improve the nation's nuclear security.

Since the 1990s, the department has touted mixed-oxide fuel as the best way to dispose of 34 tonnes of highly enriched plutonium sitting in the US stockpile, to fulfil its disarmament agreement with Russia. The technique blends plutonium with uranium so that it can be burned in conventional light-water reactors. Dozens of European reactors run on mixed-oxide fuel made from old commercial fuel, according to Matthew Bunn, a non-proliferation expert at Harvard University in Cambridge, Massachusetts.

The US plant design is actually a smaller version of the commercial French

ments require input from geologists and oceanographic experts. If there is no sedimentary rock, a country is only entitled to claim an additional 60 nautical miles out.

Satellite sensing is useless for imaging the seabed because of the overlying water, and Arctic research is further complicated by sea ice. Even when data can be gleaned by research ships, interpretation is tricky. "In many circumstances around the world when the ocean basins form and these ridges get stranded out in the middle of the ocean, it's not easy to say whether there's a greater attachment at one end of the ridge or the other," notes Parson.

In 2001, Russia became the first country to make a submission to the United Nations. It was told to supply more information, in particular about the Lomonosov Ridge, which runs under the Arctic Ocean between Russia and Canada. Russia claims this ridge is an extension of its Siberian shelf, but this is hotly contested by Canada and others. On 29 July, Russia sent two mini-submarines, Mir-1 and Mir-2 to a depth of 1.3 kilometres in the first-ever Arctic test dives to gather data to boost its claims. The subs are part of an expedition to the North Pole involving a 100-strong team of Russian scientists aboard the *Academician Fedorov*, the flagship vessel of the Federal Service for Hydrometeorology and Environment Monitoring.

Meanwhile, Canadian Prime Minister Stephen Harper has announced plans to spend around Can$7 billion (US$6.5 billion) on up to eight new reinforced Arctic patrol vessels. "Canada has a choice when it comes to defending our sovereignty in the Arctic; either we use it or we lose it," Harper said. "And make no mistake, this government intends to use it. Because Canada's Arctic is central to our identity as a northern nation. It is part of our history and it represents the tremendous potential of our future." Canada is claiming 1.75 million square kilometres of the Arctic seabed.

The Arctic is not the only contentious seabed. Countries around the world, including New Zealand, Brazil and Ireland, have already submitted claims to the United Nations. Parson is involved in the United Kingdom's four claims — one, which concerns the Bay of Biscay, will be submitted by the end of August. ■

Daniel Cressey

Melox reprocessing facility at Marcoule. Engineers have modified the design slightly to handle weaponized plutonium alloys, and they have reduced shielding because weapons-grade isotopes are less radioactive than spent commercial fuel.

The programme gained political momentum in part because Russia was planning to build a sister facility to convert an equal amount of its own plutonium surplus, according to Daryl Kimball, executive director of the Arms Control Association in Washington DC. But the Russians have backed away from the plan in recent years, Kimball says.

The US plant has also been beset by policy reviews, a lengthy licensing process and congressional opposition. It is now more than a decade behind schedule, and construction costs have ballooned by nearly 400% since 2002.

Nevertheless, department officials stand by the programme. In addition to fulfilling their disarmament obligations, the plan will be cheaper than long-term storage, Secretary of Energy Samuel Bodman told Congress.

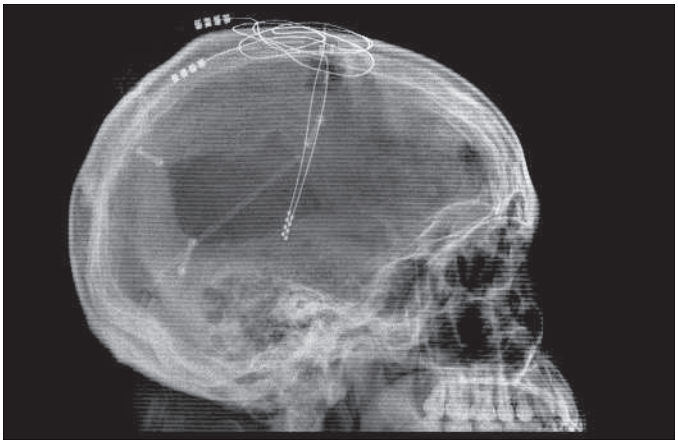
Bunn disagrees, in part because the department will be storing other plutonium anyway. "The cost of having a few extra bunkers is actually very modest," he says.

Which side Congress will take remains to be seen. The Senate is proposing roughly $390 million for construction in fiscal year 2008, but House appropriators want to give it just $168 million. A final compromise is expected to be reached later this autumn. ■

Geoff Brumfiel

NATURE|Vol 448|2 August 2007

Implant boosts activity in injured brain



Deep-brain stimulation might help trauma patients regain consciousness.

CLEVELAND CLINIC

Brain function has been improved in a patient who was in a minimally conscious state, by electrically stimulating a specific brain region with implanted electrodes. The achievement raises questions about the treatment of other patients who have been in this condition for years, the researchers say.

Patients in a minimally conscious state, often the result of severe brain trauma, show only intermittent evidence of awareness of the world around them. Typically, they are assumed to have little chance of further recovery if they show no improvement during their initial 12-month rehabilitation programme.

In the latest case study, neuroscientists describe how they implanted electrodes in the brain of a 38-year-old man who had been in a minimally conscious state for more than six years following a serious assault. By electrically stimulating a brain region called the central thalamus, they were able to help him name objects on request, make precise hand gestures, and chew food without the aid of a feeding tube (see page 600). The thalamus is involved in motor control, arousal and in relaying sensory signals — from the visual systems, for example — to the cerebral cortex, the part of the brain involved in consciousness.

Nicholas Schiff of Weill Cornell Medical College in New York, and his colleagues chose the patient because they believed his condititon was due to impairment of the arousal system, and that despite considerable damage to his cerebral cortex, many essential areas were preserved.

> "There will be a subset of patients who are responsive to this approach."

"There will be a subset of patients who are responsive to this approach," says Schiff. But he adds that patients with different brain injuries may not benefit from electrostimulation. "Not every patient in a minimally conscious state will fit this profile," Schiff says, and it is difficult for neurologists to identify those patients who will show recovery.

Nevertheless, the case shows that many patients currently seen as beyond hope of rehabilitation might benefit from the results of further research. "Severe brain injury is not an uncommon problem, and the number of people doing research on this is shockingly small," Schiff says. "It's very rare to find a programme that will take a patient in a minimally conscious state even straight out of acute care. If they don't respond in a lively enough way and can't communicate and interact with people at the bedside, they go to a nursing home directly."

"The report does not suggest that deep-brain stimulation [DBS] 'cures' the minimally conscious state," says Paul Matthews, a clinical neuroscientist at Imperial College, London. "Although based on a study of only a single patient, it suggests that DBS may be adapted to benefit at least some patients in the minimally conscious state. And it emphasizes that improvements can be made by patients even long after an injury.

"Although we do not know precisely which brain connections are important, we may expect that some specific connections must be intact for DBS to have a beneficial effect." ■

Michael Hopkin

See also News & Views, page 539.

Flatworms' starring role in stem-cell research

Tiny channels provide corridors of communication between adult stem cells and their differentiated neighbours, a study has found. The finding could be an important step towards determining how to control the development of adult stem-cell identity, but there's a catch — it was discovered in the planarian flatworm *Schmidtea mediterranea*.

The flatworms, best known for their regenerative prowess and cross-eyed charm, maintain a supply of stem cells to regrow tissue lost to anything from daily wear-and-tear to full decapitation. And, although it seems a long way to travel along the evolutionary tree, there are reasons to think planarians make a good model for human stem cells, says regeneration biologist and planarian researcher Alejandro Sánchez Alvarado at the University of Utah School of Medicine in Salt Lake City. He belives that the basic principles uncovered in planarians will translate to human cells, although the finer details may differ.

Even biologists who don't study the worms agree. "Work in planarians and other lower organisms always gives us hints regarding human stem cells," says Shinya Yamanaka, a stem-cell researcher at Kyoto University in Japan.

Biologists Néstor Oviedo and Michael Levin of the Forsyth Institute in Boston, Massachusetts, studied channels known as gap junctions that permit the passage of small molecules and ions between cells. When *S. mediterranea* was manipulated so that it did not express a key gap-junction protein, thus preventing communication between cells, it lost its regenerative capabilities and its reservoir of stem cells (N. J. Oviedo and M. Levin *Development* 134, 3121–3131; 2007).

This fits with the findings of studies in human embryonic stem cells, says Alice Pébay, a biologist at the University of Melbourne, Australia. Previous work has shown that chemically induced closure of gap junctions can slow human embryonic stem-cell growth. But, notes Pébay, Ovideo and Levin were able to use a technique to reduce expression of the gap-junction proteins — something that would be very difficult to do in human embryonic stem cells.

Nevertheless, technical challenges to planarian research remain. For example, reducing expression of an *S. mediterranea* gene is straightforward, but techniques for over-expressing planarian genes are still rudimentary, says Levin. ■

Heidi Ledford

Science embryology paper retracted after inquiry

Three out of the four authors of a *Science* paper on embryonic development have retracted it (R. M. Roberts *et al. Science* 317, **450**; 2007) after a year-long university investigation found the lead author on the original paper guilty of research misconduct.

The University of Missouri investigation was prompted by suggestions from leading embryologists that figures in the article were manipulated. It exonerated senior author Michael Roberts and two others on the paper (K. Deb *et al. Science* 311, 992–996; 2006) — but found postdoc Kaushik Deb guilty of falsifying and fabricating data.

Robert Hall, vice-chancellor for research at the university, says its findings are now under review by the federal Office of Research Integrity, which may levy sanctions against Deb. Deb's name does not appear on the retraction, and according to press reports, he fled the United States at the start of the investigation.

Coral reefs in danger from oil clean-up

J. NITO/AFP/GETTY

The chemicals used to clean up oil spills pose a far greater threat to coral reefs than the oil itself, according to an Israeli study.

Researchers at the National Institute of Oceanography in Haifa observed 100% mortality among corals treated in the laboratory with recommended concentrations of 'environmentally improved' dispersants, which break up crude oil into tiny droplets. Exposure to crude oil alone, in contrast, had no impact on the survival of the coral (*Environ. Sci. Technol.* doi:10.1021/es0704582; 2007).

Many tankers carrying crude oil pass through tropical seas rich in coral reefs. In August 2006, for example, a tanker sank near the coral-rich Guimaras Island in the central Philippines, spilling around 350,000 litres of oil into the sea (see picture). Mechanical containment and biodegrading microorganisms are often better alternatives to chemical clean-up, the authors of the study say.

Agent Orange linked to high blood pressure

The defoliant Agent Orange, used by the United States during the Vietnam War, may have caused high blood pressure in some veterans exposed to it, says the US Institute of Medicine (IOM).

Its report — the latest in a congressionally mandated series — finds "suggestive but limited" evidence of an association between exposure to Agent Orange and other herbicides used in the war and an increased chance of developing high blood pressure. Some of the herbicides used in Vietnam contained a dioxin, TCDD.

The report also reached the same conclusion regarding amyloidosis, a serious disease in which amyloid proteins are deposited in and around organs, sometimes resulting in organ failure. In previous reports, the IOM found positive associations between the herbicides and several cancers.

BETTMANN/CORBIS

Agent Orange: a legacy of ill-health.

Three-pronged genetic AIDS trial ready to go

A pioneering approach that combines gene therapy with RNA interference is to be tested on patients with AIDS.

The trial will use a gene therapy vector made from a lentivirus to shuttle in three different genetic defences against HIV. One is a short stretch of RNA designed to shut down the virus through RNA interference (see *Nature* 437, 601; 2005).

The vector will be used to treat blood-forming stem cells taken from five patients with AIDS who require bone marrow transplants to treat blood cancer. The treated stem cells will then be infused back into the patients' bodies, where it is hoped that they will combat HIV.

Molecular biologist John Rossi told the International AIDS Society meeting in Sydney, Australia, that the trial is about to begin at the City of Hope Beckman Research Institute in Duarte, California.

Network to target minorities' health

The US National Institutes of Health is to set up a network between medical schools to improve clinical research into diseases that disproportionately afflict African Americans, Hispanics and other minorities.

A three-year, US$9.5-million grant from the National Center for Research Resources at the NIH will be shared between eight medical schools that serve minority communities. As well as coordinating research into diseases such as diabetes, heart disease and AIDS, the project will seek to enrol minorities in clinical studies of new therapies.

Keith Norris, vice-president for research at Charles Drew University in Los Angeles, California, who will head the initiative, says he wants to translate leading-edge biological research into better healthcare for minorities. "We clearly have certain insights and understandings, but they haven't led to improvement in the health outcomes," he says.

MPs criticize research agency appointment

A UK parliamentary committee has criticized the appointment of engineer John Chisholm to chair the Medical Research Council, Britain's main biomedical research agency.

The House of Commons Science and Technology Committee, which oversees such appointments but cannot veto them, said it had "serious reservations" about whether Chisholm was the right person to guide the council. At a hearing on 20 June, he appeared to "lack the necessary knowledge" to discuss the MRC's relationship with other research councils, it said.

Chisholm — an engineer with a background in the computer software industry — was appointed as chair of the MRC last October, replacing another former computer-industry executive, Anthony Cleaver. A spokeswoman for the MRC said it had no immediate comment, but the government is expected to respond formally to the committee later this year.

Making room for dissent

Washington has reached an easy consensus on the need to train more scientists and engineers but, argues **David Goldston**, the United States needs to consider a broader approach to combat global competition.

Congress is an institution that tends to thrive on conventional wisdom. That may, in part, reflect the types of people it attracts — 'maverick' is not usually a term of affection on Capitol Hill — but the trait is also inherent in a body that has to reach broad consensus to get anything done. Ideas and data that don't contribute to a gathering consensus often fail to be absorbed or even heard. And notions that don't promote the agenda of any interest group may end up orphaned with no one to promote them.

The current congressional debate on competitiveness is a case in point. The conventional wisdom was encapsulated cogently in the National Academy of Sciences' 2005 report *Rising Above the Gathering Storm*. One of its main themes was that the United States is producing too few scientists and engineers because of weaknesses throughout the nation's education system.

The conclusion was familiar and sounded like common sense, and was also politically convenient. Industries that were being taken to task for sending jobs overseas or hiring immigrants could point to the study and argue that their actions reflected deficiencies in the US workforce that required long-term responses. Advocates for school systems and universities could use the report to lobby for increased funding. The recommendations in the report were fairly quickly embodied in legislation and eventually in some proposals from the Bush administration, such as the American Competitiveness Initiative. The media also picked up the theme; *Time*, for example, ran a cover story in February 2006 titled "Is America Flunking Science?".

The chorus of approval that greeted *Gathering Storm* drowned out a line of counterpoint offered by some economists who study labour markets and competitiveness. In fact, the number of US citizens and permanent residents who earn bachelor's and master's degrees in science and engineering has generally been on the rise for almost 20 years, according to data compiled by the National Science Foundation (NSF). Does this mean that the United States has enough science and engineering students? Not at all, but it's a statistic that ought to at least be part of the discussion, especially when, as economists Leonard Lynn and Harold Salzman note: "Only a third of qualified four-year college graduates in science and engineering continue in those fields." Science and engineering training can be useful in other fields, of course, but that still leaves the question of how best to attract students to science and engineering careers, if that's what the nation needs.

PARTY OF ONE

I've never heard any economist argue that the United States shouldn't invest more in improving science education or in ensuring that the country remains a leader in innovation. But many do argue for more 'nuance' than the current policy debate provides. At a recent National Academy workshop on future skill demands, some economists contended that salaries for scientists and engineers are too low. Others pointed to surveys indicating that what employers want are scientists and engineers with better 'soft' skills, such as communication. One study found that some biotech firms are trying to push more work down to the technician level, which could affect future demand for PhD scientists.

None of these ideas has the simplicity and coherence of the *Gathering Storm* recommendations, and none contributes in any obvious way to the agenda of an interest group. As a result, Congress hasn't spent much time discussing, say, what might be done to increase the demand for, rather than just the supply of, scientists and engineers, or what sort of training they should be getting in the liberal arts. Rather, in an era of unprecedented global competition, most of the policy ideas are simply recycled proposals from the competitiveness debate that occurred two decades ago, when the United States felt threatened by Japan.

Incomplete debates now could well lead to a backlash later. The last time a perceived shortage of scientists became a focus of discussion on Capitol Hill, when the Reagan administration was proposing to double the NSF's budget, the House science committee ended up investigating whether the agency had purposely manufactured a shortfall in scientists and engineers. Exaggerated promises (however sincere) about the impact of producing more scientists and engineering students may not be good for stable science funding in the long-run.

One of the few dissenting voices on science policy has been John Marburger, head of the White House Office of Science and Technology Policy. Marburger has openly questioned aspects of the prevailing consensus on increasing science funding and the output of PhD scientists and engineers — mystifying some science advocates in Washington and infuriating the rest. He has also called for the development of a 'science of science policy' to address fundamental questions such as: how do you determine the appropriate level of spending for science? In response, the NSF has started a programme to fund such research. But Marburger has gone further, arguing in a speech in May that the US government can't possibly afford to support all the new PhDs being churned out by universities.

Marburger is right to call attention to this long-standing concern about whether the population of researchers is growing too fast to be fed by the federal budget. But his antagonists on this issue can legitimately argue that the size of the budget pie isn't fixed, and another administration with different tax and foreign policies might allocate even more money to science. The matter isn't likely to be debated seriously until the message comes from another White House.

For now, Congress is likely to move ahead with the Administration's proposal to double the budgets of key physical science programmes over ten years, while providing more money than requested for other science agencies. And it is poised to pass legislation authorizing the approach called for in the *Gathering Storm*.

What needs to happen next is one of those rare, but essential periods when Congress is perplexed enough to scan the full horizon for new diagnoses and prescriptions — including some that are yet to be formulated. The next consensus will have to rely less on a reflexive turn to politically safe, time-honoured ideas if the United States is to keep enough high-value jobs to sustain its standard of living. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

BUSINESS

Swiss on a roll

Smart investors who know the drugs business well are helping to make Zurich's stock market a popular place for biotechnology firms to raise money. **Andrea Chipman** reports.

J. ARNOLD IMAGES/ALAMY

One-third of the money in Zurich's bourse relates to life science.

At a time when biotechnology companies worldwide are finding it tough to raise money, the Swiss Exchange (SWX) in Zurich is proving to be an alluring financial home.

"It's an excellent stock market that took the initiative ten years ago to build up a biotech medical segment that makes it attractive," says Ulrich Steiner, an analyst at Swiss bank Clariden Leu in Zurich, adding that smaller biotech companies find it easier to be noticed by investors and raise their profiles in Switzerland. "If you go into the London stock market, there are two to three listings every day and a lot of companies that investors could choose to invest in. In Switzerland, there's maybe one new listing every four to six weeks."

Life-sciences companies make up a large proportion of the market, accounting for one-third of total market capitalization on the SWX, or around €245 billion (US$335 billion), which includes some major international pharmaceutical companies. That's the largest concentration of its kind in Europe, according to the *Swiss Biotech Report 2007*, a survey published in April by Ernst & Young, the Swiss government and other parties.

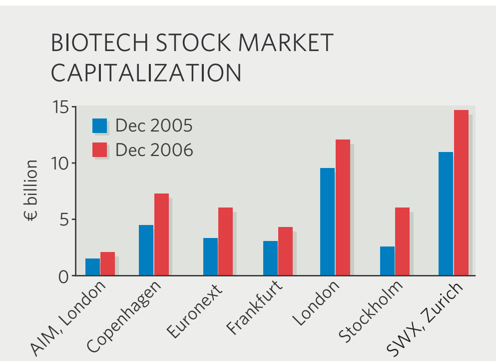
When it comes to the drugs business, Switzerland has lots of advantages. Aside from more than 200 established biotech companies, and some of the world's largest pharmaceutical firms, such as Novartis and Roche, it has an extensive biotechnology infrastructure. Well-funded universities, federal and cantonal government support for fledgling companies and a host of organizations and initiatives that provide coaching, promotion, advice and patent-application help for start-ups have nutured the sector. Identifiable biotech clusters have developed in four regions of the country.

And many of the factors that have made Switzerland a friendly place for biotech start-ups — a favourable tax environment and a multilingual population and workforce — have also attracted investors and analysts with a highly specialized biotechnology knowledge, market-watchers say. Although most of the life-sciences companies on the SWX are home-grown, foreign ones are increasingly taking notice. In 2006, two out of three initial public offerings (IPOs) in biotech companies were from over the border, including Italian companies BioXell and Newron — spin-offs of Roche and Pharmacia & Upjohn, respectively. The total volume of companies debuting on the market remains modest compared with those of London and New York, however.

Investors are encouraged by the recent performance of Zurich's sector-specific indices, such as SXI Life Sciences and SXI Bio+Medtech, which have tended to outperform similar measures in Frankfurt and London (see graph).

And according to Jürg Zürcher, an industry analyst at Ernst & Young in Basel, the SWX raised more money from biotech IPOs in 2006 than any other stock exchange in Europe.

Last December, for example, Newron of Milan, Italy — a company specializing in treatments for disorders of the central nervous system and pain management — raised €75 million by floating on the SWX. According to Stefan Weber, the company's chief financial officer, it scoured the world's stock markets beforehand: it excluded Nasdaq after deciding that it would need to base a major part of its operations in the United States to gain the attention of US investors. Weber says that didn't make sense, adding that the Sarbanes–Oxley financial reporting regulations and the difficulties of getting analysts to cover any company worth less than $500 million also kept the company away from Nasdaq.



SWISS BIOTECH. RPT. 2007

Newron also discounted most of the major European stock markets — including its home market of Milan — because of the generally unimpressive performance of life-sciences companies there both before and after share offerings. In Milan, the shortage of specialized indices, investors, investment banks and analyst coverage counted against it. Turning to the London markets, Newron weighed the higher market capitalization and global sales of the London Stock Exchange against the lack of specialized sector coverage. When it came to London's Alternative Investment Market (AIM), Newron was sceptical about its abilities to compete for coverage and market attention with lots of other small companies that Weber characterizes as "low-quality".

"We thought we had a premium business case," he says. "We wanted to have the best environment possible." That meant Switzerland, where Newron was able to do a fully international flotation, with 70% of investors coming from outside the country.

Switzerland's markets are dominated by well-known private banks and a large number of wealthy individual retail investors. That fits better with the biotechnology industry than the fast-paced trading cultures in London and New York do, analysts say.

"The specific criteria of life-science stocks are different from, say, those of a car manufacturer," explains Hanns-Peter Wiese, a partner at Munich-based Global Life Science Ventures. "You need a dedicated clientele to buy into those stocks who understand them, and don't just expect to make a quick buck. Swiss investors seem to understand, seem to appreciate, and seem to make money." ■

GETTY

The other greenhouse effect

Rising carbon dioxide levels should increase crop yields. But what if their effect on the nutritional value of our food is less benign, asks **Ned Stafford**.

When you step into a commercial greenhouse, the chances are you are stepping into the future. To plants, carbon dioxide is food, and greenhouse operators, knowing this, use it to fatten them up. While today's atmosphere contains about 380 parts per million of carbon dioxide, commercial greenhouses often contain carbon dioxide concentrations of twice that or more — the sort of concentration that we might expect in the open air at the end of the century.

The fact that plants thrive in environments with high levels of CO_2 — all other things being equal — seems to offer a silver lining to the otherwise dark clouds of climate change. Many crop scientists believe that this carbon dioxide fertilization effect will go at least some way towards offsetting the losses in yield to be expected as a result of the higher temperatures, flooding, drought and rising sea levels that the CO_2 greenhouse effect will bring. The fact that many horticulturalists already choose higher CO_2 environments for their work underlines this assumption.

But some are not so sure. They are sounding alarm bells about potential negative impacts; bigger yields, they say, are not always better. Their worry is that the nutritional value of crops could suffer regardless of overall abundance. These researchers point to the known negative effects that increased carbon dioxide concentrations have on the protein content of crops. They also worry about subtler effects that might be felt in everything from the micronutrient properties of soya beans to the ability of wheat to be baked into bread. They do not all share the same level of concern, but they do all agree on one thing. Compared to the amount that is being spent on climate research, the amount being spent on understanding the agricultural effects of higher CO_2 levels is woefully inadequate.

Steven Adams, head of strategic and applied research on the physiology of protected crops at the University of Warwick's Horticulture Research International in the UK, reflects the relatively relaxed attitude of most crop scientists. He acknowledges that hardly any research has been done on the effect of high–CO_2 in greenhouses on food quality and concedes that a drop in nutritional value is possible. "If it affects photosynthesis and yields, it could do," he says. "But I would have thought the impact is relatively small."

A more complex view comes from Bruce Kimball, a soil physicist with the United States Department of Agricultural in Maricopa, Arizona — a pioneer in high-CO_2 plant research. When he started out in the 1970s, available technology for high-CO_2 research was limited mainly to greenhouses and open-top chambers, both of which offer questionable results. Plants grown in such systems are not exposed to environmental variables such as wind and normal variations in temperature and humidity.

In the mid 1980s, Kimball was a driving force in development of a system dubbed FACE (free air carbon dioxide enrichment), which simulates natural field conditions. A FACE system includes a ring of equipment up

to 30 meters in diameter encircling the plants that are the subject of the research. The vertical pipes in this ring emit CO_2; sensors measure wind speed, wind direction and carbon dioxide levels. A computer uses this data to keep the CO_2 within the ring at the target level by releasing it from specific upwind pipes as required. It's not a perfect simulacrum of a high-CO_2 world—for one thing, the gas is pumped in only while there's daylight for the plants to photosynthesize in. But it's the best so far.

After his decades of work Kimball agrees that higher CO_2 concentrations can lead to lower nutritional quality of crops. But he sees nothing in his FACE studies — or those of others — to cause alarm. Based on current knowledge, he says, the net effect of increasing CO_2 is a good one: "As far as crops go, I think higher CO_2 is a definite benefit. Yes, a little less nutrition than before, but we get more food."

The plants almost always deliver higher yields than controls, with more sugar and starch in their leaves. They also take up less nitrogen from the soil, because they are making less protein. A lot of the protein in leaves is involved in assimilating CO_2 into sugars. At higher CO_2 levels that becomes easier; less protein is needed, and so less protein is made. The major exception is in the 'C4 plants', which are better at photosynthesis in less favourable conditions and so are less susceptible to the effects of changing CO_2 levels. C4 crops include maize, sorghum and sugar cane.

But while Kimball thinks that, in general, the gains in yield are the most important thing, he is not blind to the drop in protein levels. Talking of data from a cotton-leaf experiment, he finds himself struck by the size of the effect: "That is a big change," he says, wondering aloud what that might mean for lettuce, other plant leaves and grasslands. "Think of all the livestock that only eat leaves." Grass-grazing livestock in the 550 parts per million CO_2 world that is likely, though not inevitable, by 2050 might be getting significantly less protein from their forage.

And it's not only livestock that eat plants — there's the rest of the ecosystem too. Kimball conducted side experiments in the 1990s when growing crops in his FACE systems. He describes the effects of a diet of high–CO_2 cotton leaves on beet armyworms. "What we found is that their growth and reproductive capability was reduced," he says.

But Kimball believes that the protein levels can be lifted by increasing the amount of nitrogen supplied to the plants. "The farmer in the future would have to be sure and apply ample fertilizers to keep protein quality up," he says. And he thinks crop scientists and plant breeders will be able to rectify most other potential problems arising from high-CO_2 levels. But he believes more money is needed to fund high–CO_2 food research. Operating FACE systems is expensive — according to Keith Lewin of Brookhaven National Laboratory, US, a forest-based FACE system costs about $1.5 million a year to run even before you include the costs of research — and Kimball's last major field experiments were in 1999. "After that, funding dried up," he says. "I think we need to do more experiments at much higher CO_2 concentrations," he goes on, noting that nearly all CO_2 plant research up to the present has been done at 550 parts per million, not at any of the higher levels that are conceivable in the second half of the century.

Feedback from wind sensors helped Bruce Kimball control the release of CO_2 in a FACE study on wheat.

J. DYKINGA

"If we see this change in just five years, what will happen in 50 years?" - Stephen Long

But while Kimball is confident that increasing nitrogen levels could help retain higher protein levels in plants, others are not so sure. Arnold Bloom, a plant biologist at the University of California, Davis, believes the reduction in nitrogen content seen in plants in a high CO_2 environment is not just due to a lowered need for protein in leaves. He thinks it is because of a decreased ability to take-up nitrates from the soil. If this is the case, then adding nitrogen fertilizer would be less effective in future, which could have implications both for protein content and for overall yields (see 'Diminishing returns').

The question of whether adding nitrogen can offset the effects of increased carbon dioxide remains open, according to Hans-Joachim Weigel, director of the Institute of Agroecology, part of the Federal Agricultural Research Centre in Braunschweig, Germany. For some crops you have to add "enormous amounts of nitrogen", amounts that would be unfeasible in terms of cost and unacceptable in terms of environmental damage.

Weigel's FACE research on barley, wheat and sugar beet in the past six years leaves no doubt in his mind that higher CO_2 levels in coming decades will have a significant impact on crop quality. "We should be concerned about it, but not in a panic about it," he says. Herbert Wieser, head of cereal proteins at the German Research Centre for Food Chemistry in Garching, has provided some of the grounds for that concern by milling Weigel's winter wheat into white bread flour and analyzing the flour for protein content.

Wieser found that high-CO_2 wheat, grown with a normal amount of nitrogen fertilizer, produced white bread flour with 7.8 grams of total protein per 100 grams of flour, 14% less than the 9.1 grams of protein in flour from wheat grown at normal CO_2 levels. When the wheat was grown with half the usual fertilizer, the protein level was 6.7 grams in the control, and 6.1 grams in the high-CO_2 group.

The total protein content is not the only thing that changes — so does the type of protein. Gluten proteins are used as a nitrogen store in wheat, and in high-CO_2 conditions this store was lowered more than the overall protein level, dropping by 18%. The high-molecular-weight subunits which are particularly important for dough and bread quality fell by 23%. Wieser concludes that high-CO_2 growing conditions cause "a serious impairment of wheat baking quality".

This goes some way towards vindicating Andreas Fangmeier, professor of plant ecology and ecotoxicology at the University of Hohenheim in Germany, who was quoted in a university-issued press release earlier this

Diminishing returns

Higher levels of carbon dioxide stimulate overall crop yields if other conditions remain suitable. But increases in crop yields may not be as great as once thought.

Studies from the 1970s and 1980s offered the optimistic possibility that CO_2 fertilization could offset weather-related losses due to climate change. But newer data offer less hope, says Stephen Long of the University of Illinois. In soya bean studies that he has compared, those in closed systems show yields up an average of 28–30% in an atmosphere with 550 parts per million CO_2, compared with an increase of only 13–15% in studies done in open-field conditions using FACE (free air carbon dioxide enrichment) technology.

In general, he says, FACE studies indicate yields increasing only half as much as they did in studies done in closed chambers. Overall, too little has been done to assess the impact of climate change on crops and the lack of data makes it hard to be precise. "We are not in a good position right now to predict our future," Long says. "We need to start preparing or we are going to have problems."

Pramod Aggarwal, head of environmental sciences at the Indian Agricultural Research Institute in New Delhi and a co-author of the food, fibre and forest-products section of the Intergovernmental Panel on Climate Change's latest report, says the idea that CO_2 fertilization will offset weather-related losses "is a Western point of view" that ignores much of the developing and underdeveloped world. He argues that high CO_2 levels are beneficial only when crops are adequately fertilized, irrigated and protected from pests, so food crops in tropical climates will derive less, if any, yield benefit from increased CO_2 concentrations.

Take rice for example. According to Reiner Wassmann, coordinator of the Rice and Climate Change Consortium established by the Philippines-based International Rice Research Institute in 2006, whereas rice yields increase under higher CO_2 levels, the probable higher temperatures could depress yields. But the research infrastructure to study future conditions is inadequate. Only two large FACE systems are dedicated to rice, one in Japan and the other in China: there is no research being done in tropical nations. Wassmann thinks improved temperature tolerance might be achievable, but it requires more research. "Adaptation of rice production to climate change will require substantial funds to support vigorous and concerted efforts by national and international research institutions," he says.

Aggarwal is more concerned about the global quantity of food in coming decades than about quality. "The food supply must be maintained," he says. "That is the primary goal. We need to have something to eat, even if it is low quality. To have nothing is a big problem." **N.S.**

year warning that by 2050, CO_2 concentrations could make French fries poisonous, beer foamless and wheat flour unbakeable. Unfortunately, Fangmeier did not actually bake bread, brew beer or fry chips from crops harvested from his high-CO_2 test fields. Chemical analysis of his crops merely indicates "the potential" for the problems foreseen — the press release, he admits, was "an exaggeration".

Despite his lapse into hype, though, Fangmeier refuses to back down from the underlying sentiment that higher atmospheric CO_2 levels will damage crop quality in a number of ways. For example, he notes that in one of his high-CO_2 field tests, Vitamin C in potato tubers dropped by 50–60%. He can't explain the drop, but says it is just one of many indications that high-CO_2 levels will have a much larger impact on food quality than many currently believe. Weigel says the database for speculating about such things is currently too small.

Another researcher convinced that protein levels are not the whole story is Irakli Loladze, now a mathematician at the University of Nebraska in Lincoln. In 2002, while doing postdoctoral work in mathematical biology at Princeton University in New Jersey, Loladze published an opinion article (I. Loladze *Trends Ecol. Evol.* **17,** 457–461; 2002) highlighting the risks of changed plant composition in a high-CO_2 world and focusing on micronutrients such as iron, iodine, copper and manganese. Scouring the "surprisingly scant" literature he found that, on average, the concentrations of all the micronutrients he looked at decreased. The plants put more effort into storing up carbohydrates, and the resultant "carbohydrate dilution" reduced the proportion of metabolically important trace elements such as chromium, selenium and zinc.

Loladze believes that with current atmospheric CO_2 levels a third higher than pre-industrial levels, plants have already changed in this way, and that they will inevitably change more. He's concerned that this could put hundreds of millions at risk of the "hidden hunger" of micronutrient malnutrition. However, he has been unable to attract significant funding for further research into the area.

UIUC-ACES-ITCS

Soil cores measure how much water plants in a elevated CO_2 environment take from the soil.

Peter Curtis, professor of ecology at Ohio State University in Columbus, Ohio, says Loladze's 2002 paper identified "a clear gap in our understanding of plant responses to elevated CO_2: how this global change in plant nutrition will affect plant tissue micronutrient status". Even subtle changes in micronutrient status could affect both human health and the wider ecosystem.

Some research has bolstered Loladze's argument, at least in part. Stephen Long, a crop scientist at the University of Illinois at Urbana-Champaign, says that soya beans grown in his FACE fields have shown drops in calcium and zinc levels of 10–20%, sugars and starches have been up 50%. The drop in calcium might be particularly noteworthy, as soya beans are used to make substitute dairy products.

His team also found effects passing up the food chain. For example, western corn rootworms feeding on high-CO_2 soya bean leaves live longer and produce twice as many young as those feeding on normal soya bean leaves. "They lay eggs in the soya bean fields in late summer so that they are ready to infect the corn crop that will be planted in the next year," Long says, referring to the standard practice in the region of rotating corn and soya beans annually. They have also measured changes in the microbial community within the FACE rings. "We don't know whether this is good or bad," he admits. "But if we see this change in just five years, what will happen in 50 years?" ■

Ned Stafford is a freelance writer based in Germany.

NATURE|Vol 448|2 August 2007

TRIAL AND ERROR

The ethics committees that oversee research done in humans have been attacked from all sides. **Heidi Ledford** recounts the struggle to come up with alternatives.

Fourteen years of treating people with tuberculosis has taught physician William Burman what to expect when a patient walks through his door. Tuberculosis is not typically a disease of the well-heeled. Many patients in the United States are foreign born. English is their second language. Fewer than half have completed a high-school education, and many have spent time in jails or homeless shelters.

So when Burman, of the University of Colorado in Denver, joined in two studies run by the Tuberculosis Trials Consortium, he knew that the consent forms needed to cater to people with an eighth-grade reading level (comprehensible to an educated 13-year-old). The trials involved multiple institutions, and the forms were sent to 39 institutional review boards (IRBs) — committees designed to determine whether a proposed experiment is ethically sound. The final approvals came in 346 days later, but what the IRBs sent back, Burman found disturbing.

"The consent forms were longer. The language was more complex," Burman says. "And errors were inserted at a surprising frequency." In one case, a potential negative side effect of the treatment had been accidentally edited out. Burman responded to the problem as any researcher would: he studied it. He had an independent panel review the changes. The reviewers found that 85% of the changes did not affect the meaning of the consent forms, but that the average reading level had jumped from that of an eighth grader to that of a twelfth grader (around 17 years old)[1]. His results confirmed something he'd suspected for some time. "I started to think about what was happening and it just seemed like the system was flawed." It was time to change the system.

Burman is not alone. In the 40 years since their birth (see 'Time for ethics'), IRBs, also known as research ethics committees, have faced criticism from all sides. They're too slow, or too hasty, overprotective, or they flout basic safety. They're bureaucratic, wasteful and unavoidable. So, what are researchers to do? The will for change exists, says Sarah Greene, a researcher at the Group Health Center for Health Studies in Seattle, Washington. But recent attempts to fix the system have struggled to gain a foothold.

Obstacle course

In many countries, a complex network of local ethics committees handles the approval of research on humans. This focus on local resources allows committees to account for specific laws or cultural concerns in a particular region. But it leads to problems in multicentre trials, such as Burman's, which are becoming more frequent. When IRBs were first founded, multicentre trials were almost unheard of. A 1998 report[2] from the inspector-general of the US Department of Health and Human Services in Washington DC stated that a rise in the number of multicentre studies was throwing the system into crisis. And a recent analysis[3] showed that five of 20 trials seeking IRB approval reported significant delays as a result of IRB negotiations. Seventeen noted inconsistencies both in IRBs' review process and in their recommendations. In one case, negotiations between 65 IRBs delayed the study by a year.

> **"A national IRB could fail to recognize different needs of different communities."**
> — Lainie Ross

More frighteningly, the cumbersome system could even endanger the health of the studies' participants. The higher the hurdles — and the more unfair they seem — the less inclined researchers will be to jump them. "It's slow and frustrating to researchers," says Ezekiel Emanuel, chair of the US Department of Bioethics at the National Institutes of Health in Bethesda, Maryland. Researchers have reported that they are more likely to violate the regulations set by ethics committee if they feel that they or their application have been mishandled[4].

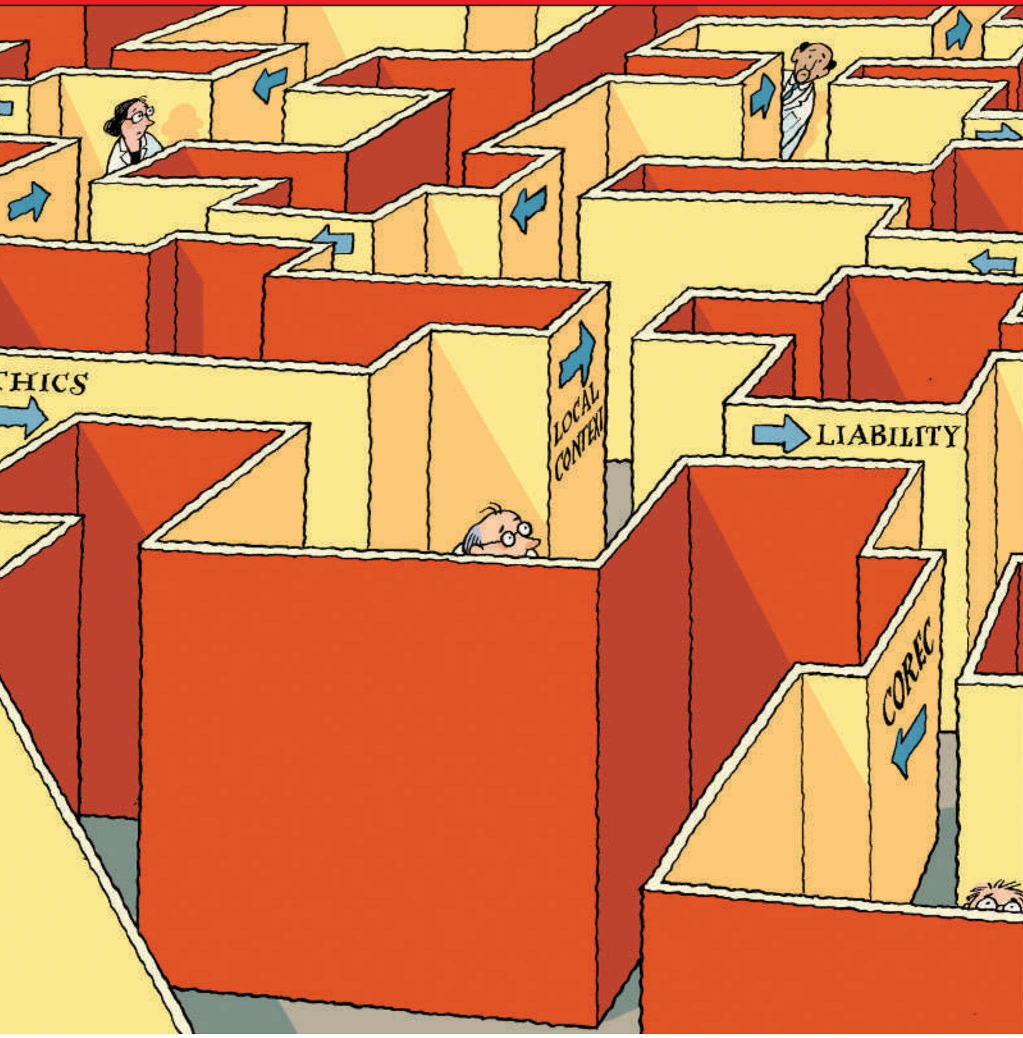
And even though IRBs are made up mostly of volunteers, they are expensive to run. In 2002, the median cost of running an IRB, taking into account the time spent by IRB members, was $742,000; the maximum was over $4 million[5]. So, for every protocol they assess, they charge a fee to cover support staff, facilities, and outside consulting. These fees are typically pulled from grants as part of the institutional overhead, or as direct charges to commercial sponsors, and they average just over US$1000 (ref. 5).

A proposed solution to the copious problems with IRBs is outsourcing, especially for multicentre studies, to some form of centralized review. That movement has met with resistance from those who say that local review provides valuable local context. But Burman counters that local context had little bearing on the changes in his consent forms. In his trials, only 1.5% of the tweaks to the consent forms were made to account for local context[1].

Risky process

But that's not enough to rule out the importance of local review argues David Wynes, vice-president for research administration at Emory University in Atlanta, Georgia. "I agree that the vast majority of changes are editorial, but I think there's a value in an institution having a process for identifying when local context is an issue," he says. "You might have to review a hundred protocols before you can see the value of local context. Is it OK if only 1% of the time you put subjects at risk?"

Burman argues that expertise with a specific patient group or disease should trump local context from detached review boards. "The local IRBs don't know the patients I take care of, because if they did, the last thing they



D. PARKINS

would do is increase the length of the consent form and make the language more complex," he says. Instead, Burman and his colleagues have worked to create a designated panel at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, which keeps track of disease epidemiology in the United States, to review all tuberculosis studies.

Top trumps

But a centralized system will work only if the local boards are not allowed to overrule the decision of the central board, says Emanuel. It's a lesson, he adds, that the United Kingdom has had to learn the hard way.

In 1997, the United Kingdom created a system of regional review boards in which trials needed approval from just one board to proceed. In 2000, the system was brought under the auspices of the Central Office of Research Ethics Committees (COREC), based in London. The trouble was, local ethics committees refused to surrender control, and instead of expediting review, COREC had created a new layer of bureaucracy.

"Researchers were very upset with the way things were going," says Emma Cave, a lecturer at the Leeds School of Law, UK. "They thought the regulations were making the United Kingdom a bad place for research." In April, Britain dissolved COREC in favour of the new National Research Ethics Service, and changed the regulations to restrict the ability of local ethics committees to change the protocol approved by the national office.

In 2001, the US National Cancer Institute (NCI) in Bethesda, Maryland, launched a similar experiment — a central review board to review all NCI-funded research on humans. Local review boards retained the power to do a full review, but could opt instead for an expedited review in which they merely adjust for local context. (The NCI formed a similar review board for paediatric studies in 2004.)

> "Is it OK if only 1% of the time you put subjects at risk?"
> — David Wynes

The project immediately ran into trouble. The central board spent too much effort on scientifically reviewing proposals that had already been reviewed by the granting arm of NCI, says Richard Schilsky, chairman of Cancer and Leukemia Group B, an NCI-sponsored cancer clinical trials group. And at first, few local IRBs were willing to cede control to the central review board. "The concept is good," says Schilsky, "but the devil has been in the details of the implementation."

Since 2001, the number of participating institutions has climbed to 300. More than half of those have accepted the reviews of NCI's central board; the remainder are still developing ways to incorporate the central boards review into their own review process.. But Schilsky says that only about 20% of the institutions involved in his clinical-trials group — the largest in the United States — have signed on. The result was similar to what happened in Britain, adding to the bureaucracy. Schilsky estimates that the system has added two to three months to the time it takes to activate a new study.

Lainie Ross, a paediatrician and member of the IRB at the University of Chicago Medical School, says that she is opposed to surrendering local control. "A national IRB could fail to recognize different needs of different communities. I'm not just going to accept someone else's word for it."

Without fail, IRB members interviewed by *Nature* who were opposed to ceding control to a centralized board cited concerns about patient safety as their main reason. But Emanuel, who has also served on an IRB, says that there's another cause for concern. "There are no good data suggesting that there are local factors that are ethically relevant," says Emanuel. "It's really liability that's driving this."

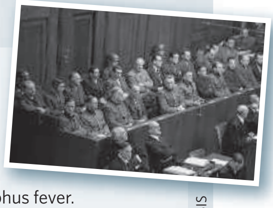
Vulnerable populations

Liability is a thorny issue for local IRBs contemplating handing over control to NCI's central IRB, says Wynes. If a participant in a clinical trial felt that he or she were unjustly harmed during the course of the research, they could not hold the NCI legally responsible because it is a branch of the federal government. That leaves the local IRB legally vulnerable, says Clint Hermes, general council at St. Jude Children's Research Hospital in Memphis, Tennessee.

It is rare, but IRBs and even individual IRB members have been sued in the past. Bioethicist Arthur Caplan at the University of Pennsylvania in Philadelphia says that he has served on two IRBs that were sued but still thinks it's important to have a mechanism in place to hold negligent IRBs accountable.

Meanwhile, a profitable industry in private, commercial IRBs has sprung up. Although commercial IRBs can provide a sense of security by assuming legal liability, the institution doing the experiment will bear ultimate responsibility. But partial indemnity seems sufficient to comfort many researchers: commercial IRBs serve hundreds of companies, hospitals and research institutions. In 2005, the consulting firm Deloitte named Chesapeake Research Review of Columbia, Maryland

TIME FOR ETHICS



BETTMANN/CORBIS

1947
The Nuremberg Code governing human research is released as a response to the Nazi experiments of large-scale sterilization and vaccine testing on prisoners intentionally infected with typhus fever.

1963
Doctors at the Jewish Chronic Disease Hospital in New York City inject cancer cells into patients as part of a study of transplant rejection. Patients were not told what was in the injections.

1964
The World Medical Association issues ethical principles governing human research called the Declaration of Helsinki. These later become the basis for ethics committees around the world.

1966
Britain establishes its first ethics committee. The *New England Journal of Medicine* publishes a paper citing 22 violations of research ethics at major research institutions in the United States.

CORBIS SYGMA

1972
Authorities terminate a 40-year experiment in which treatment was withheld from men with syphilis in Tuskegee, Alabama. The men were told that they were receiving free treatment.

1974
In response to the Tuskegee scandal, America founds the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in Washington DC — its first office dedicated to protection of human subjects.

1997
Controversy erupts over US-funded trials in Thailand testing the effect of zidovudine (AZT) on HIV transmission from pregnant women to their children. Women in the placebo group received no treatment.

J. KEYSER/TIME LIFE PICTURES/GETTY IMAGES

KRT/NEWSCOM

1999
18-year-old Jesse Gelsinger dies from complications of a gene-therapy experiment at the University of Pennsylvania in Philadelphia, launching a barrage of criticism about gene therapy and human research.

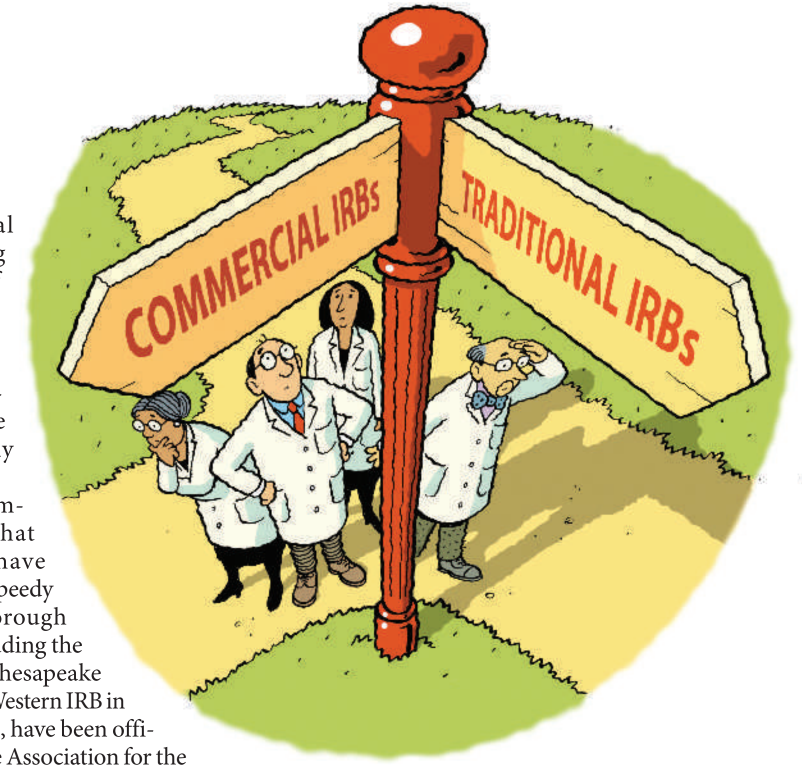
2000
Britain establishes the Central Office of Research Ethics Committees in London to review multicentre trials. America refuses to sign a revision of the Declaration of Helsinki that limits the use of placebo-controlled trials.

2001
The US National Cancer Institute launches its central review board, to evaluate all human research funded by the institute. The board aims to standardize and expedite ethics review.

2007
In response to criticisms about COREC, the UK forms the National Research Ethics Service and restricts the ability of local review boards to demand changes in experimental protocols.

— a commercial IRB and consulting service — as one of the fastest-growing technology companies in North America. The firm increased its revenues by 244% in five years, to reach nearly $5.5 million in 2004.

Proponents of commercial IRBs say that larger companies have good reputations for speedy turnaround and thorough reviews. Several, including the two largest players, Chesapeake Research Review and Western IRB in Olympia, Washington, have been officially accredited by the Association for the Accreditation of Human Research Protection Programs in Washington DC. Such societies provide a stamp of approval for IRBs, providing oversight and standardization to the field.



D. PARKINS

Financial gain

Still, others worry about the potential conflict of interest inherent to commercial IRBs, who could benefit financially from pleasing their customers and passing protocols with minimal fuss. "I'm a little cautious about this drive towards commercial IRBs," says Richard Bianco, associate vice-president for regulatory affairs at the University of Minnesota in Minneapolis, and a 15-year IRB veteran. Nevertheless, Bianco and other critics acknowledge that local IRBs also have a conflict of interest — clinical trials can bring in serious cash and prestige to the institutions they serve, and IRB members that are also scientists at the institution may feel pressured to approve a trial. "I'm somewhat surprised that no one has ever pushed to reform the private side of IRBs," says Caplan. "It's growing like crazy. Industry hires them because they're fast and efficient. It doesn't mean that they're right."

> **"Industry hires private IRBs because they're fast and efficient. It doesn't mean that they're right."**
> — Arthur Caplan

Bianco also says that his colleagues have been under "intense pressure" by industry collaborators to relinquish control to commercial IRBs. Some trial sponsors, he says, even issue ultimatums: use the commercial IRB that we recommend or don't participate in the trial. "That was pressure," says Bianco. "But I've been around a long time. You come to know what to ignore." None of those threats ever came to fruition, he says.

For others, initial scepticism of commercial IRBs has given way to acceptance. "When I first came across independent IRBs, I questioned them, too," says Wynes. "But I've taken the time to get to know how they operate, and my comfort level has changed." In November 2005, while Wynes was still at the University of Iowa in Iowa City, he helped the university to switch to outsourcing industry-sponsored trials to Western IRB. Prices vary, but outsourcing to industry can cost twice as much as processing the application in-house, although commercial IRBs typically boast a quicker turnaround time. Whereas many commercial IRBs aim to review applications within a week of their receipt, non-commercial IRBs may meet only once a month.

Some see commercial IRBs as a stop-gap measure in lieu of real regulatory change. But despite the roadblocks, substantial change is inevitable, says Emanuel. The lingering problem, he adds, is that it will probably take a new scandal to push reform to the top of the agenda. "I think we're just one accident away, but it will still take the accident," he says. "In my opinion, that's the sad fact." ■

Heidi Ledford writes for *Nature* from Boston, Massachusetts.

1. Burman, W. *et al. Control. Clin. Trials* **24,** 245–255 (2003).
2. Office of the Inspector General, Department of Health and Human Services. Institutional review boards: a time for reform. (US Government Printing Office, Washington, DC, 1998).
3. Greene, S. M. & Geiger, A. M. *J. Clin. Epidemiol.* **59,** 784–790 (2006).
4. Keith-Spiegel, P. & Koocher, G. P. *Ethics Behav.* **15,** 339–349 (2005).
5. Sugarman, J. *et al. N. Engl. J. Med.* **352,** 1825–1827 (2005).

See Editorial, page 511.

Profit or perish even worse than publish or perish

SIR — Michael Alvarez, in his Recruiters article 'System-ready scientists' (*Nature* **447,** 612; 2007) advises young scientists to consider how they can fit into and enhance the labour force, even though this course might seem dehumanizing to them. He advises universities to gather information from government and industry about their 'design specifications' for incoming talent, to help them develop graduate training programmes that produce suitably qualified trainees.

I believe such an approach is a backwards step. Many scientists study and do research out of irrepressible curiosity and sheer love for the subject. It is true that the US academic funding system created a glut of postdoctoral researchers who have had to find alternatives to conventional career paths, but 'publish or perish' is a relatively harmless dictum compared with 'profit or perish'.

Rather than helping universities strip the creativity, intuition and joy of discovery from scientists while they train, government and industry should be willing to spend money, within their pragmatic frameworks, to acknowledge that scientific research is a fundamentally creative pursuit.

Edward Kiegle
Moonshine Mushrooms, PO Box 927, 251 Putnam Hill Road, Chester, Vermont 05143, USA

Interdisciplinary research could pull cash into science

SIR — James A. Smith and Gemma E. Carey, in Correspondence (*Nature* **447,** 638–639), address the need for supportive environments if interdisciplinary research goals are to be achieved. Among other points, they mention that assessments of such research quality are made in single-subject committees in New Zealand, Australia and the United Kingdom.

I, and I am sure others, made similar points to a UK Parliamentary Science and Technology Committee general research review in 1999 (see www.parliament.the-stationery-office.co.uk/pa/cm199900/cmselect/cmsctech/196/196ap02.htm).

Several useful developments occurred as a result. Generic funding initiatives were set up in genomics, e-science and information technology, where inter-disciplinary investigators and teams could apply to tackle broad research objectives. Discipline-hopping awards were established to encourage researchers to be mobile in their research interests. The UK Research Councils took a unified 'road show' around universities and institutes to take input from the research community. Articles were published in UK Research Councils newsletters, describing research-interface developments. Input was added to research assessments, to flag across-the-disciplines relevance.

These efforts did not succeed in achieving a unified science and engineering research council, with a single charter and unified management structure. However, I believe that the government made efforts to respond to criticisms and, as a result, put the United Kingdom in a more flexible position.

There remains another big issue on the block, as Philip Strange pointed out in Correspondence (*Nature* **448,** 22; 2007): the ratio of grant proposals made to those funded can be between three and five to one. This is a very arbitrary measure of research quality. Seeing excellent proposals, agreed by referees and my own experience, finally get rejected, shows all too clearly that the 'unfunded ratio' is deeply flawed as a metric, at least at this level of underfunding. Perhaps interdisciplinary research could be the latest big idea for bringing new money into the United Kingdom's national research enterprise?

John R. Helliwell
School of Chemistry, University of Manchester, Manchester M13 9PL, UK

The 'hundred surnames' of China run into thousands

SIR — I am writing to point out one misconception that may arise from your Editorial 'Asia on the rise' (*Nature* **447,** 885; 2007). Although the most common names in China are known as 'the hundred surnames', there are many others in widespread use.

A recent survey by the Chinese Academy of Sciences found 4,100 surnames, of which 129 covered 87% of the population (Y. Yuan *Chinese Nat. Geog.* **2,** 38–39; 2007, or see http://english.people.com.cn/200601/11/eng20060111_234647.html). This leaves nearly 200 million people using the other surnames.

The mistaken idea that only 100 surnames are used probably comes from a famous tenth-century booklet that is literally translated as *A Hundred Family Names*. However, in Chinese, 'a hundred' often just means a large number — in fact, the booklet itself included 504 surnames. The *Grand Dictionary of Chinese Surnames* (Y. Yuan and R. Du, Education and Science Press, Beijing, 1996) included 11,969 surnames, and 24,000 have been recorded in the past.

The practice of taking the mother's surname as well as the father's has often been proposed, but I doubt whether the Chinese government will take any measures to enforce it in the near future.

Dafeng Hui
Auburn University School of Forestry and Wildlife Sciences, 602 Duncan Drive, Auburn, Alabama 36849, USA

University speaks up in dispute over cancer centre

SIR — As vice-president of public affairs at Arizona State University (ASU), I would like to give a different account of recent events from that put forward in Correspondence by Carl Djerassi ('Not so sunny view of the events in Arizona' *Nature* **447,** 1052; 2007).

At ASU, as at most American research universities, positions funded by external grants are completely dependent on that funding. Robert Pettit, former director of the university's Cancer Research Institute (CRI), attempted to secure funding to continue his research, submitting proposals to both the National Cancer Institute and the National Oceanic and Atmospheric Association, but these were not selected by those agencies from among the many competing proposals. That is the sole reason why 22 staff and seven students working on Pettit's research team (not 60 individuals as Djerassi indicates) had to be reassigned or released.

The affected individuals were given a 30-day notice of termination and offered outplacement support. Of the 22 permanent staff, seven were placed in comparable positions at ASU and two chose to retire.

Pettit remains a Regents professor in the department of chemistry and biochemistry with tenure at ASU. The non-renewal of his year-to-year administrative appointment as director of CRI was based on findings by the university provost, among other things, that Pettit had contravened policies of the Arizona Board of Regents and ASU.

Virgil Renzulli
Arizona State University, PO Box 872503, Tempe, Arizona 85287-2503, USA

Climate information helps homeowners make choices

SIR — You raise the question in your News story 'Website homes in on climate hazards' (*Nature* **447,** 360–361; 2007) of whether homeowners can make good use of the climate risk information provided by our company. We have had tens of thousands of visitors to our site, and thousands of requests for reports, indicating that homeowners want this information to make their own property risk assessment, given the potential risks to their safety, retirement equity and quality of life. We have heard clearly that people feel empowered by having this information to assess their risk for themselves, rather than waiting for the possibility that an insurance company's underwriting changes will affect their options and their property value.

David Purcell
Climate Appraisal Services LLC, 25 Abbey Road, Easton, Connecticut 06612, USA

BOOKS & ARTS

Social climbers

Does a baboon's success in social situations depend solely on learning the rules of the game?

Baboon Metaphysics: The Evolution of a Social Mind
by Dorothy L. Cheney & Robert M. Seyfarth
University of Chicago Press: 2007. 358 pp. $27.50

Asif A. Ghazanfar

M. BOTZEK/ZEFA/CORBIS

Understanding how baboons interact could throw light on the origins of the human mind.

There are few guiltier pleasures than watching reality television such as *Big Brother*, in which young people interact with each other in confined spaces. Viewers relish seeing each individual trying to push forward their own agenda through alliances and disagreements, sex and friendships. How did we become both agents and voyeurs of such status-striving?

In *Baboon Metaphysics*, Dorothy Cheney and Robert Seyfarth explain that our social reflexes evolved from our group-living primate ancestors. They explore what sort of intelligence is required to navigate the intricate social landscape that baboons live in. Is it based on a complex calculation, a system of innate rules that are applied to specific contexts? Or is it based on simple, implicit rules governed solely by learned associations?

The book's title comes from a line that Charles Darwin jotted down in his 1838 *Notebook M*: "He who understands baboon would do more towards metaphysics than Locke." The quotation reflects the tension between two philosophical schools of thought on the origins of knowledge. One, espoused by John Locke, suggests that the mind acts simply to associate events that have been joined together by proximity and repetition. The other, represented by Immanuel Kant, suggests that perceptions exist *a priori* — that is, the mind is not a blank slate — but require experience for their expression. Darwin, a witness to the stereotyped behaviours of numerous animals, could not abide Locke's view and sought an explanation of the mind that combined the roles of innate tendencies and experience. This tension pervades this wonderful book on the social intelligence of non-human primates and what they might tell us about the evolution of the human mind.

Few are in a better position to address the question than Cheney and Seyfarth. They have spent many years observing and conducting behavioural experiments on vervet monkeys in Amboseli National Park, Kenya — the subject of their earlier book, *How Monkeys See the World* (University of Chicago Press, 1990; see *Nature* **350,** 565; 1991 for review), and now work on baboons in the Okavango Delta, Botswana. They share this experience with us in a lively and engaging manner.

First, they lure the reader in with wonderful anecdotes, such as Ahla, the goat-herding baboon, who spontaneously recognized the relationships among her goats, and compulsively reunited any lost baby goats with their mothers. Or there's the orphaned baboon who, when separated from his group, cleverly spent a few days under the protection of vigilant groups of impala, and later vervets, before a jubilant reunion with his fellow baboons.

Next they provide experimental evidence to tease apart what baboons really know about social relationships and how they use this information to get ahead in the world. For example, playing back a sequence of calls that mimics a dispute between individuals reveals that eavesdropping females are acutely aware of who is fighting and whether it involves their family. More transient social relationships, such as sexual dalliances, are also closely monitored. Playing back a recording of a male's grunts with a female's copulation call reveals that hopeful bachelors are acutely aware of another male's consortship, or when a female is making a cuckold of him or when a consortship has ended.

Cheney and Seyfarth argue that this complex social knowledge cannot be the result of simple associative learning alone, particularly because it does not necessarily result in immediate rewards or benefits. They posit, for example, that baboons learn about hierarchies through observing the close associations between certain individuals. The authors suggest that the number of dyads and triads to learn is too vast, and that putative metrics, such as rates of aggression, do not unambiguously specify the nature of a relationship because they occur with similar frequencies both within and between different family groups. Furthermore, baboons belong to many different social classes concurrently — for example, a female can be a member of matrilineal group, a friend of a high-ranking male, and/or a friend of other females outside her kin — and class membership is liable to change.

What, in addition to associative learning, is necessary to explain baboon behaviour? Cheney and Seyfarth suggest that evolution selected individuals who are predisposed to recognize other individuals' ranks and social relationships to form rule-governed classes. They propose that this predisposition is innate and similar to the human predisposition to learn language. Unfortunately, this is not a very satisfying argument as it leaves much of a baboon's behaviour to some mysterious

innate mechanism. One possibility not mentioned by the authors is that the baboons' social knowledge is based on statistical learning, in which relationships can be implicitly learned through the increased probabilities that certain dyads or triads are seen together or that certain sequences of calls are heard with greater frequency than others. This mechanism is different from typical associative learning in that it is rapid, does not require reward and can be used to generate rules. Under this scheme, any explanation of baboon social intelligence would be completely interdependent with the structure of the current social group and the interactions therein.

We reflexively attribute minds like ours to non-human agents. Cheney and Seyfarth are acutely aware of this and, with their clever field experiments and careful observations, they address important questions regarding the evolution of social cognition without succumbing to the almost irresistible temptation of anthropomorphizing. Their enthusiasm is obvious, and their knowledge is vast and expressed with great clarity. All this makes *Baboon Metaphysics* a captivating read. It will get you thinking — and maybe spur you to travel to Africa to see it all for yourself. ■

Asif A. Ghazanfar is an assistant professor at the Neuroscience Institute, Department of Psychology, Princeton University, Princeton, NJ 08540, USA.

The atomic peacemaker

Einstein on Politics: His Private Thoughts and Public Stands on Nationalism, War, Peace, and the Bomb
edited by David E. Rowe and Robert Schulmann
Princeton University Press: 2007. 560 pp. $29.95, £18.95

Yaron Ezrahi

As a German Jew who rose to be the most celebrated scientist since Newton, a pacifist triggered by the rise of Hitler to recommend the development of the atomic bomb, a cosmopolite driven by the fate of his people to support a Jewish nation state, or as an émigré to America who supported socialist ideas in the time of McCarthyism, Einstein was often at the centre of clashing ideologies. A solitary individual who became trapped by the limelight of the world stage, Einstein was reluctantly forced to become an activist. Thus, *Einstein on Politics* is a goldmine for readers interested in Einstein as an engaged intellectual of his era.

Editors David E. Rowe and Robert Schulmann have done an excellent job of collecting, thematically assembling and historically contextualizing Einstein's private letters and public statements on the great political issues of his time. The book is also a fascinating record of Einstein's private thoughts and public stance on the reception of the relativity revolution. Included here are his reaction to the virulent anti-Semitic, anti-relativity German scientists, his tortured relations with the Prussian Academy of Sciences after the rise of the Nazis and his later expressed identification with Galileo for his struggle "to overcome the anthropocentric and mythical thinking of his contemporaries and to lead them back to an objective and causal attitude towards the cosmos".

Einstein's statement in 1921 that "my Zionism does not preclude cosmopolitan views" could serve as the motto for the vast sections in this book documenting his engagement with the fate of Jews. His efforts to reconcile his cosmopolitanism and Zionism found particularly clear expression in his deep engagement with the founding of the Hebrew University of Jerusalem in 1925. Einstein regarded a Hebrew university as a vital part of a Jewish renaissance in Palestine and also as a necessary place for gifted Jewish youths barred by anti-Semitism from many European universities. Whereas the inclusion of 'Hebrew' in the name of the university implied a commitment to creating a Jewish home, Einstein had great faith in the mission of the university as an international academic institution. In a statement in March 1925, Einstein insisted that "Jewish nationalism is today a necessity" and that, together with other educational institutions, the Hebrew University should regard it as one of its "noblest tasks to keep our people free from nationalistic obscurantism and aggressive intolerance".

This book also includes fascinating documentation of Einstein's private and public responses to the rise of Nazism, in the course of which he forged an influential exemplar of the morally engaged twentieth-century intellectual. Collaborating with other prominent activists such as Romaine Rolland, Sigmund Freud and Bertrand Russell, he continually expressed his hope that the principles taught by great Germans such as Kant and Goethe would some day "prevail in public life and the general consciousness". This goal required that scientists and other intellectuals would assume public responsibility as advocates of tolerance, rational discourse, non-violence and other humanistic values. Provoked by the accusation from the Prussian Academy that his public statements against fascism constituted "atrocity-mongering against the German people", Einstein insisted on the moral responsibility of intellectuals to speak out against violent nationalism. When urged by the German physicist Max Von Laue to exercise some restraint, he responded "Does not such restraint signify a lack of responsibility? Where would we be had men like Giordano Bruno, Spinoza, Voltaire and Humboldt thought and behaved in such a fashion?"

C.R. MACAULEY

Einstein takes up the sword against fascism in this 1933 cartoon from the *Brooklyn Eagle*.

Einstein's impact on the relations between science, politics and freedom, however, transcends his record as a public intellectual. Ironically, the unintended wider cultural legacy of his physics worked against his commitment to democratic values and his faith in the mission of scientists to publicly combat violence and irrational politics.

In a letter to Rolland in August 1917, Einstein insisted that "only facts can dissuade the majority of the misled from their delusion". But Einstein's concept of facts, as expressed in his exchange with the French philosopher Henri Bergson, was rather esoteric. Failing to appreciate the importance of common-sense realism as the basis of democratic public discourse, he did not seem to anticipate that the shift from newtonian to einsteinian physics would widen the gap between authoritative scientific knowledge and lay opinion. His liberal-democratic commitment was contradicted by his view that "naive realism", the belief that "things 'are' as they are perceived by us through our senses", was a "plebian illusion". Deeply concerned about the turning of the public into a herd in the country of Kant and Goethe, he also failed to see that the public in democratic

societies is not exactly moved by rational arguments free from rhetoric and theatricality.

In the final analysis, when Einstein thought or spoke about democracy, he seemed to have focused on conditions for the freedom and creativity of the individual rather than the group. Hence, much more important than all the things he said and wrote was his personal example as a scientist. By demonstrating the powers of one person's mind to revolutionize the view of the cosmos, Einstein reaffirmed the value of free individual thinking in resisting group-mind and coerced opinion — and therefore sustained the possibility that the majority could be wrong. ■

Yaron Ezrahi is a professor of political science at the Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel.

Nervous systems made simple

An Introduction to Nervous Systems
by Ralph Greenspan
Cold Spring Harbor Laboratory Press: 2007. 200 pp. $45, £26.99

Eve Marder

Until I saw pieces of giant squid in a fish market in Chile, I somehow hadn't appreciated that giant squids were really giant. Ralph Greenspan's *An Introduction to Nervous Systems* will similarly enlighten many of its readers on the wonders to be found through the study of invertebrate nervous systems and the behaviours they control. Furthermore, it is an eloquent mixture of fundamental neuroscience and evolutionary biology. Greenspan uses a variety of invertebrate animals to describe the fundamental processes of nervous-system function and by placing the basics of neuronal signalling in their functional contexts, he casts this information in evolutionary terms.

The savvy and sophisticated neuroscientist will find themselves learning about evolution; the evolutionary biologist will be led through a remarkably clear exposition of ion channels and action potentials, the fundamental elements of nervous activity. Smart high-school students will find an accessible and engaging account that reveals the magic and mysteries of nervous–system function in a wide range of animals. The book will be especially useful as a text for university-level courses wishing to provide students without previous knowledge of neuroscience a broad context with which to understand how nervous systems generate behaviour.

Greenspan manages to be a scholar without being a pedant. Each chapter, which starts with an apt quotation from the classics, has a specific focus, and examples are chosen with finesse. Greenspan introduces the fundamentals of ion-channel structure and function using the swimming behaviour of the unicellular *Paramecium*. He continues with the escape response of jellyfish and includes a remarkable tidbit of cellular electrophysiology that I was unaware of: jellyfish neurons that generate either a low-amplitude calcium-dependent or a high-amplitude sodium-dependent action potential to trigger a slow or rapid behavioural response, respectively.

Greenspan introduces neuromodulation in the context of the behavioural sensitization of the gill-withdrawal response of the sea slug *Aplysia californica*. This is the famous preparation used as a paradigm to study simple forms of learning, and here is my first quibble. I am, perhaps wrongly, sceptical of the generality of the results that are used to argue that some species of mollusc fail to show behavioural sensitization. I wonder whether assaying a different behaviour might reveal sensitization in the species that don't show it using the methods so effective for *A. californica*.

Circadian rhythms, flight, navigation, decision-making, courtship and mating in *Drosophila* are used to introduce the structure and function of networks important for complex behaviours. These sections of the book are presented with enough detail to do justice to the interplay of physiological processes important for an animal to function in the real world. Greenspan's scrupulous respect for understanding an animal's nervous system in terms of its real-world behaviour is apparent in his discussion of honeybee learning, where he relates the probable connection between the timing of stimuli in honeybee learning and the nature of the bees' foraging behaviour.

Greenspan clearly intended this book to be readable by and accessible to anyone. He succeeds admirably and it is an unadulterated delight to read. However, I confess to a modicum of regret that to avoid interrupting the flow, the references are placed at the back of the book, without citing which experimental result is found in a given reference. This will require readers to dig a little (not such a bad thing) to find some of the papers that they might wish to look at more carefully.

In addition to its use as a textbook, this book would make a wonderful present for your lawyer, banker, doctor or musician friend who retains a genuine curiosity about how animals interact with the world. Alternatively, you can take it to the beach or mountains this summer and use it to remind yourself to watch all kinds of creatures in their natural settings — while they and their environments still exist on our planet. ■

Eve Marder is the Victor and Gwendolyn Beinfield Professor of Neuroscience at Brandeis University, Waltham, Massachusetts 02454-9110, USA.

J. HUEMOELLER/PHOTOLIBRARY.COM

Even the simplest animal nervous systems, such as those of jellyfish, can spring suprises.

SCIENCE IN CULTURE

C. SAVILLE/APEX

Sculptor Peter Randall-Page with the finished *Seed* before it is transported to the Eden Project.

Consulting nature's pattern-book

Peter Randall-Page's massive granite sculpture for the Eden Project emulates natural form.

Long before scientists applied their analytical skills to underlying regularities of form and process in nature, artists intuitively exploited a repertoire of common patterns that recur across organic and inorganic forms. Some favourites, such as spirals and waves, feature as motifs in an astonishingly wide range of cultures over the ages.

Contemporary artists who are drawn to these patterns begin from similarly intuitive starting points, drawn into the magic of form through our ingrained fascination with order and disorder in natural process. Nowadays the artist, if he or she chooses, can look to a large body of research and writing by scientists who have examined the mechanisms behind appearance.

For their research an artist can turn to distinguished books of popular science such as Philip Ball's *The Self-Made Tapestry*, John Barrow's *The Artful Universe*, and the various explorations of the mathematics of complexity by Ian Stewart. And, of course, they can always consult the grandfather of geometrical morphologists, D'Arcy Wentworth Thompson, whose 1917 book, *On Growth and Form*, remains an all-time classic of science writing.

Peter Randall-Page has established himself as one of the foremost explorers of this territory, as a sculptor who has consistently deployed the geometry of natural growth. He defines his art as founded on "nature's underlying pattern-book of forms".

Seed, the massive granite sculpture recently installed in the education building at the Eden Project, an environmental centre in Cornwall, UK, is a triumphant realization of his vision. Weighing 62 tonnes and standing more than 4 metres tall, its great granite mass is enclosed in the tight embrace of a pod-like chamber at the centre of the building, where it is subject to the infinitely varied caress of natural light.

The curvaceous egg shape of *Seed* was determined by Randall-Page's sense of the instinctive rightness of its contours in relation to size, mass and material. A set of nodes, systematically graded in their diameter and protrusion, spiral around the form. The relationship to nature is one of principles and resonances, rather than imitation. We think of a pine cone or the arrangement of bud primordia on a shoot.

Randall-Page explains how this phyllotaxis was achieved. "I initially worked with Iain Cant, an expert in computer three-dimensional modelling. Together we plotted the spiral phyllotaxis pattern onto a virtual three-dimensional form that was then printed out as a three-dimensional rapid prototype model. I had hoped that we would be able to project the virtual image onto the stone itself in order to draw the pattern, but in the event this method proved not to be sufficiently accurate, although it was useful for establishing the arrangement of circles on the very top of the sculpture.

"Eventually, I plotted the pattern of around 1,800 nodes directly onto the surface of the stone using a ruler and compass. I drew two primary spirals traversing the form in opposite directions to represent the two dominant alignments of circles. Using horizontal bands, I was then able to divide each circumference into numerical divisions of two consecutive Fibonacci numbers, in this case 21 and 34. Joining these points created two families of opposing spirals whose intersection represented the centre of each node."

The resulting sculpture is more than simply an adornment to the building that houses it. Randall-Page played an important role in how architect Jolyon Brewis at Grimshaw determined the spiral geometries of the 'sunflower seed' roof that embraces the lofty peak of *Seed*. The sculpture and building share their natural engineering. They sit in the Eden clay pit as beautiful products of human creativity — what Leonardo da Vinci called "a second nature in the world". ■

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.

NEWS & VIEWS

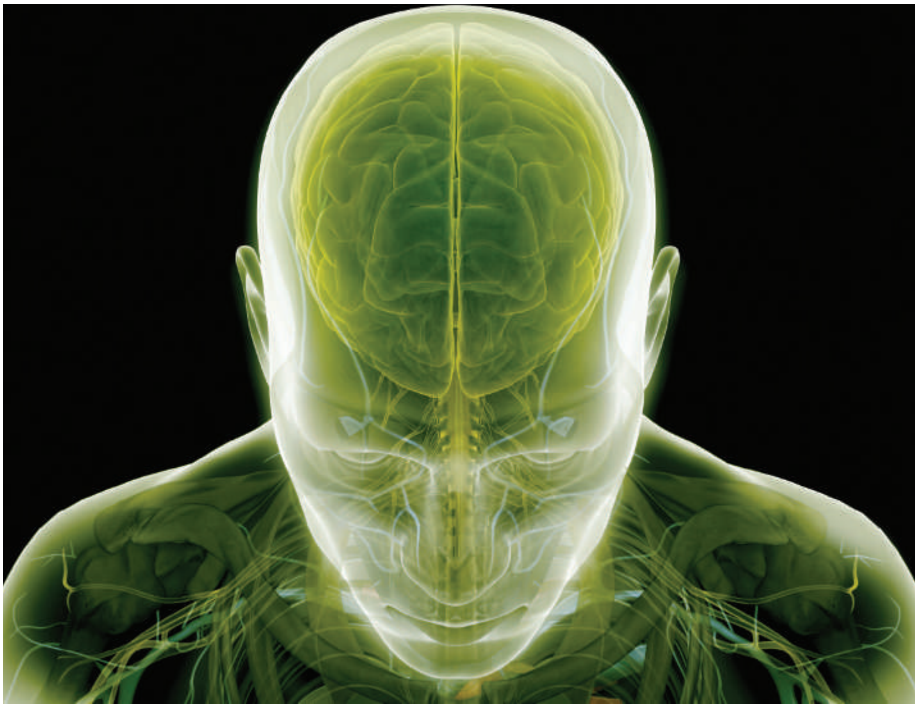
NEUROLOGY

An awakening

Michael N. Shadlen and Roozbeh Kiani

Neuroscientists and engineers are developing ways to help patients overcome paralysis and stroke. But what about mental function itself? Can medical intervention restore consciousness?

Jean-Paul Sartre wrote[1]: "In one sense choice is possible, but what is not possible is not to choose." To the neurologist, however, gaining consciousness is a decision of the unconscious brain to make choices. Philosophers and scientists may argue about the definition of consciousness[2,3], but neurologists have little trouble identifying its absence. Now, physicians are beginning to understand how it can be restored in some patients with severe brain damage. A case report by Schiff *et al.* (page 600 of this issue[4]) raises hope in this area, and sheds light on the neurobiological underpinnings of consciousness. Schiff and his colleagues treated a patient who had been in a 'minimally conscious state' (Box 1, overleaf) for several years after a serious brain injury.



3D4MEDICAL.COM/GETTY

Sadly, the vast majority of coma patients do not recover consciousness. The prognosis is determined by the type of injury to the brain, its extent, and the findings from serial neurological examinations[5]. For example, a trained neurologist can predict with near certainty that meaningful recovery will not occur for many patients who remain in a coma for days after a cardiac arrest, in which the brain is deprived of blood flow and oxygen. For other patients, however, the outcome is less certain.

Even after severe brain injury, some patients retain enough of the cerebral cortex to raise hopes that some degree of organized mental function might one day recover. Indeed, some show intermittent signs that are clearly distinguishable from coma, despite an overall level of function that is effectively unresponsive. For these patients, we do not have reliable indicators of prognosis, and we lack treatments that might help the brain restore consciousness.

But advances in basic neuroscience are beginning to reveal the brain systems that are responsible for monitoring and sustaining engagement with the world around us. A key component is the thalamus, which lies between the brainstem and the cerebral hemispheres, and forms the gateway to the brain's cortex.

The thalamus is organized as a set of nuclei. The best understood of these nuclei are those containing the neurons that relay information from the eyes, ears and skin to the appropriate sensory cortex. But much of the thalamus is poorly understood. Anatomical studies in non-human primates have identified a class of thalamic neuron that might operate more generally in activating cortical networks[6]. These neurons, which stain positively for the calcium-binding protein calbindin, are found in all thalamic nuclei. Although we know little about the physiological properties of these calbindin-positive cells, they tend to exhibit a different pattern of connections with the cortex compared with the relay cells. Their axons terminate more broadly both across cortical areas and in layers that the relay cells miss. These calbindin-positive cells comprise a large percentage of the intralaminar nuclei of the thalamus — nuclei that have long been thought to have a role in arousal.

Schiff *et al.*[4] hypothesized that their patient might express a minimal level of consciousness because of a primary impairment of the arousal system itself. The patient had suffered irreparable damage to much of the cerebral cortex, but many essential areas were preserved. By stimulating the intralaminar nuclei, the authors hoped to switch on the undamaged areas of cortex. Neurologists and neurosurgeons have previously used electrodes to monitor brain activity in patients with epilepsy and to stimulate deep-brain regions in the treatment of severe Parkinson's disease. Because the brain itself lacks sensory receptors (after all, it is normally protected by a cranium), these electrodes cause no discomfort. This insight, and extensive experience with stimulation electrodes in animal experiments[7], made the procedure feasible and relatively safe. Such considerations probably helped to guide the complex ethical debate preceding this experimental trial on a human patient.

The results were dramatic. Within 48 hours of the surgery to place the electrodes, the patient, who had remained in a minimally conscious state for 6 years, demonstrated increased arousal and sustained eye-opening, as well as rapid bilateral head-turning

to voices. Schiff and colleagues allowed 50 days of post-operative recovery before activating the stimulating electrodes again to ascertain that the stimulation by the implanted electrodes — not some unknown aspect of the surgery — was the source of the improvements.

An 18-week 'titration' phase, which involved finding the most effective patterns of stimulation, followed. During this phase, the previously non-verbal patient became capable of naming objects and using objects with his hands — for example, bringing a cup to his mouth. Moreover, he could swallow food and take meals by mouth, meaning he was no longer dependent on a gastrostomy tube.

After the titration phase, the authors turned the stimulation on and off systematically for a better assessment of its effects. The highest state of arousal and the ability to perform functional limb movements were associated with periods of stimulation. But compared with the pre-surgical period, many functions remained improved even without stimulation, indicating carry-over of the stimulation effects to no-stimulation periods.

These results raise hope for some patients with brain injury, but there are two caveats. First, this is a single case report; a carefully controlled clinical trial is needed to answer the remaining questions. What is the spectrum of brain injury that will benefit from thalamic stimulation? What are the clinical, anatomical and functional (neuroimaging) predictors of improvement?

Second, not all patients with disorders of consciousness will benefit from thalamic stimulation. This patient had shown clear signs of interactive behaviour and preservation of many of the important cortical structures before the surgery. Thalamic stimulation presumably increased both the level and consistency of activity in the preserved cortical structures, leading to arousal and behavioural improvements. Such stimulation would not benefit patients who have already lost the critical cortical structures; the condition of persistent vegetative state that often follows deprivation of the brain of oxygen and blood flow (hypoxia and ischaemia) falls into this category.

But besides the hope that this study furnishes for some patients, the observations of Schiff *et al.* may provide clues about the neurobiological underpinnings of consciousness. Cognitive neuroscience is beginning to expose the architecture of information processing that is directed towards goals and actions; we refer to this as an 'intentional framework', but it flies under various banners, including 'affordances' and 'embodied cognition'[8–10].

In essence, the brain does not process information in the abstract but instead consults information acquired through the senses and in memory insofar as it bears on the decisions made about potential actions and strategies. Our brains allow us to decide among possible options — that is, how and in what context to engage with the world around us. The brain makes many such decisions unconsciously. Indeed, the decision to engage at all is, in effect, an unconscious decision to be conscious. Thus, the brain of the sleeping mother queries the environment for the cry of her newborn. We suspect that the normal unconscious brain monitors the environment for cues that prompt it to decide whether to awaken and engage. This mechanism may be disrupted in various disorders of consciousness, including the minimally conscious state, hypersomnolence, concussion, abulia (lack of will) and possibly severe depression.

Previous theories of consciousness have relied on a central executive and magical physiological phenomena (for example, synchronized reverberations) to elevate the subconscious functions of the brain to consciousness. However, viewed as a decision to engage, consciousness can instead be studied in the same framework as other types of decision and the allocation of attention[11]. Rather than a central executive, there seems to be a network of brain regions that organize the resting state and maintain overall orientation towards context[12,13]. It is quite possible that they make decisions about whether or not to engage and in what way. They do what Sartre considered impossible: they choose whether to choose or not.

How these networks relate to intralaminar nuclei or the matrix of calbindin-positive thalamic neurons is another question. However, the idea that these areas need to be turned on for consciousness leads us to wonder whether stimulation of the intralaminar thalamus in this patient[4] worked through the activation of calbindin-positive neurons and these circuits. If so, then the work of Schiff *et al.*[4] could point the way towards interventions that are more refined than stimulation with electrodes.

These are loose strands, but neuroscience is beginning to stitch together the neurologist's and the philosopher's ideas of consciousness. It is wonderful to see the fruits of this research help one patient. We hope that further studies will advance our knowledge and help more patients. Meanwhile, our advice is: treat your blood pressure, wear a helmet and read the existentialists. ■

Michael N. Shadlen and Roozbeh Kiani are in the Howard Hughes Medical Institute, Departments of Physiology & Biophysics, and Neurology, University of Washington Medical School, Box 357290, Seattle, Washington 98195-7290, USA.
e-mails: shadlen@u.washington.edu; roozbeh@u.washington.edu

Box 1 | Disorders of consciousness

To be awake is to be in a state of engaging with the environment. To have agency is to interrogate this environment with some goal or purpose. This capacity to engage and interrogate seems to go awry in various disorders that affect cognition. And nowhere is this incapacitation more patent than in coma and related conditions.

Coma. The patient seems to be asleep and cannot be awoken. There is no spontaneous organized behaviour, not even pushing away of an irritant, and there is no evidence of any awareness of sensory cues — no response beyond reflexes mediated by the brainstem and the spinal cord. Most patients in coma do not recover meaningful neurological function, but many do progress to states that are clearly distinguishable from coma, as described below.

Persistent vegetative state (PVS). This is similar to coma in all respects except that, at times, the patient does not seem to be asleep. The eyes may be open, and spontaneous, non-purposeful, roving eye movements occur. PVS is probably the result of a return of some of the functions that would govern the sleep–wake cycle, albeit in the absence of a functioning cerebral cortex[14]. As a result, it is only the brainstem and perhaps a few islands of dysfunctional (or disconnected) cortex that regain function.

Minimally conscious state (MCS). In contrast to PVS, patients show occasional signs of arousal and organized behaviour. Nevertheless, for the most part, there is a profound deficit in consciousness. Indeed, functional brain-imaging studies[15] indicate that parts of the cortex may be able to function even when the patient seems to be unconscious. This observation and the differential prognosis of MCS and PVS call for more careful classification of patients, which may currently be biased towards PVS[16]. **M.N.S. & R.K.**

1. Sartre, J.-P. *Existentialism and Human Emotions* (Carol, New York, 1984).
2. Crick, F. & Koch, C. *Nature Neurosci.* **6,** 119–126 (2003).
3. Block, N. *Trends Cogn. Sci.* **9,** 46–52 (2005).
4. Schiff, N. D. *et al. Nature* **448,** 600–603 (2007).
5. Levy, D. E. *et al. J. Am. Med. Assoc.* **253,** 1420–1426 (1985).
6. Jones, E. G. *Trends Neurosci.* **24,** 595–601 (2001).
7. Shirvalkar, P., Seth, M., Schiff, N. D. & Herrera, D. G. *Proc. Natl Acad. Sci. USA* **103,** 17007–17012 (2006).
8. Merleau-Ponty, M. *Phenomenology of Perception* (Routledge & Kegan Paul, London, 1962).
9. Clark, A. *Being There: Putting Brain, Body, and World Together Again* (MIT Press, Cambridge, Massachusetts, 1997).
10. Cisek, P. *Phil. Trans. R. Soc. B* published online 11 Apr 2007 doi:10.1098/rstb.2007.2054 (2007).
11. Gold, J. I. & Shadlen, M. N. *Annu. Rev. Neurosci.* **30,** 535–574 (2007).
12. Dosenbach, N. U. F. *et al. Proc. Natl Acad. Sci. USA* **104,** 11073–11078 (2007).
13. Fox, M. D. *et al. Proc. Natl Acad. Sci. USA* **102,** 9673–9678 (2005).
14. Kinney, H. C. & Samuels, M. A. *J. Neuropathol. Exp. Neurol.* **53,** 548–558 (1994).
15. Laureys, S., Owen, A. M. & Schiff, N. D. *Lancet Neurol.* **3,** 537–546 (2004).
16. Bernat, J. L. *Neurology* **58,** 337–338 (2002).

CLIMATE CHANGE

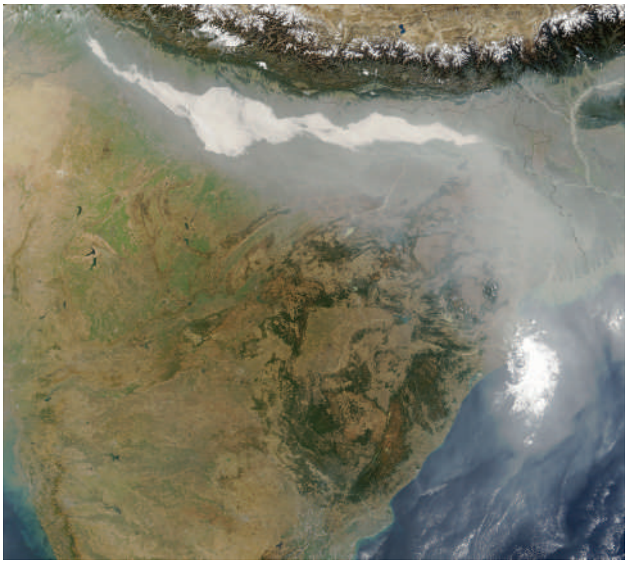
Aerosols heat up

Peter Pilewskie

Solid particles suspended in the atmosphere have long played second fiddle to greenhouse gases as agents of climate change. A study of atmospheric heating over the Indian Ocean could provoke a rethink.

In the fourth assessment report of the Intergovernmental Panel on Climate Change (IPCC), released earlier this year, the effect on climate of aerosols — small, suspended particles of varying composition, size and shape — since the start of the industrial era was estimated to be about 20% of that of greenhouse gases[1]. Aerosols are thought to have a cooling effect on the atmosphere, and therefore to have mitigated some of the expected global warming over this period.

This is, however, a highly uncertain conclusion, in part because the total amount and vertical distribution of solar radiation that is absorbed by aerosol particles is imperfectly known. In this issue, Ramanathan *et al.* (page 575)[2] report that the aerosol clouds above large regions of Asia actually cause as much warming as greenhouse gases — in contradiction, at first glance, to the notion of aerosol particles as a cooling agent.



NASA/GODDARD SPACE FLIGHT CENTER/J. SCHMALTZ

Figure 1 | Hazy days. **Smog drifts down India's populous Ganges valley and out into the Bay of Bengal. This is the source of 'atmospheric brown clouds' over the Indian Ocean, and the climatic effect of its constituent aerosol particles is investigated by Ramanathan and colleagues[2].**

The effect of aerosols and greenhouse gases on the climate is usually described as 'radiative forcing', an often-used but frequently misunderstood term. Simply put, it is the perturbation to the net rate of radiative energy flow in the atmosphere caused by atmospheric constituents such as clouds, aerosol particles or gases. Aerosol particles, for example, scatter solar radiation in all directions, so more solar radiation should be reflected back into space when aerosols are present. Thus, aerosol particles have a net cooling effect, and their radiative forcing is negative.

Radiative forcing is usually defined at the top of the atmosphere[3], or more specifically, at the top of the troposphere (this is the lowest 8–12 km of the atmosphere, which contains most of the aerosol particles, clouds and water vapour). Thus calculated, radiative forcing is a useful parameter in climate modelling both because the globally averaged temperature at Earth's surface responds linearly to forcing, and because a model's response can be tested and compared in the absence of feedback mechanisms and independently of the forcing agent[3].

But for aerosol particles that absorb radiation (for example, soot produced by combustion), top-of-atmosphere forcing tells only part of the story. When solar radiation is absorbed, it heats the atmospheric layer in which the particles reside, even though the net effect for the entire atmospheric column may still be cooling.

Ramanathan *et al.*[2] obtained the crucial data missing from model- and satellite-derived forcing estimates that are needed to determine the extent of this heating of the lower atmosphere. They deployed three small unmanned aerial vehicles (UAVs) carrying instruments to measure aerosol-particle concentrations, aerosol absorption and solar radiation near the remote Indian Ocean island of Hanimaadhoo, in the northern Maldives. Such UAVs are promising tools for airborne science, particularly for radiation studies in which simultaneous, vertically aligned observations are needed to examine the profile of radiative energy absorption. The small UAVs used by the authors were developed 15 years ago[4] to measure parameters such as air temperature and pressure, and are used in much the same way that weather balloons are used to gather the data for forecast models. Unlike balloons, UAVs can be directed to probe specific regions of interest.

For their 'Maldives autonomous campaign' (MAC)[2], the authors let their UAVs fly in coincident vertical formations between 0.5 and 3 km above sea level throughout March 2006. This is the middle of the region's annual dry season and is typified by low-level pollution moving from the Asian continent to the northern Indian Ocean.

Two distinct sets of results were identified during the campaign. The first half of March was dominated by a marine airmass, whereas in the second half air arrived from south Asia that had higher aerosol-particle concentrations and aerosol absorption. During this second, pollution-dominated period, atmospheric heating increased by more than 50%, a finding validated by model studies using data acquired at a climate observatory on Hanimaadhoo. Most of the heating increase occurred in the visible part of the electromagnetic spectrum, confirming that the elemental carbon in soot was primarily responsible.

Although the direct observations occurred over a relatively small area and over a short period, they are representative of widespread plumes of pollution, known as atmospheric brown clouds, that blanket much of south Asia during the dry season (Fig. 1). To account for the variability of these plumes in time and space, Ramanathan *et al.* relied on satellites and a network of surface observations from an earlier study[5] covering 2000–03. The regional variation in total aerosol concentration and absorption between pollution-dominated and pollution-free cases throughout this period was in close agreement with the detailed MAC study.

In addition, Ramanathan *et al.* determined how aerosol particles are distributed with altitude in the atmospheric brown clouds using data from the recently launched cloud-aerosol lidar and infrared pathfinder satellite (CALIPSO). CALIPSO revealed that the highest concentration of pollution is at altitudes between 1 and 3 km, again confirming the MAC flight measurements. The authors therefore infer that the MAC results are representative of the larger-scale regional forcing that affects global climate.

To assess the longer-term picture, Ramanathan and colleagues simulated heating caused by atmospheric brown clouds during the period 1950–2000 by imposing the estimated forcing during 2000–03 on the US National Center for Atmospheric Research's general circulation model CCM3. They determined that warming due to greenhouse gas during that period was between 0.5 and 0.8 °C, about the same as that from the atmospheric brown clouds. The trend in total warming is an increase of about 0.25 °C per decade, twice the rate of warming at the surface, and it has been confirmed by satellite-based microwave measurements. These trends have substantial implications for the elevated region of the Himalaya, where observed warming of 0.15–0.3 °C during the past several decades has led to the rapid reduction of glacier mass.

These findings[2] might seem to contradict the general notion of aerosol particles as cooling agents in the global climate system, but they also emphasize the importance of resolving how aerosol radiative forcing varies with altitude. Even though the global view depends on satellite- and ground-based observation networks, intensive studies from airborne platforms, which have become more versatile and efficient with the advent of UAVs, are essential for advancing climate research. Small fleets of UAVs could be used in other regions of the globe to examine the climatic impacts of, for example, regional pollution outflow from urban areas, airborne dust and effects from cloud–aerosol interactions.

A further perspective is that from space. In late 2008, NASA is scheduled to launch the Glory mission, which will fly an aerosol polarimeter sensor with advanced capabilities of retrieving aerosol particle information, including composition and absorption, not attainable from current space-based instruments[6]. Together with ground- and air-based observations, Glory should help us to assess the contribution of aerosols to global climate. ■

Peter Pilewskie is at the Laboratory of Atmospheric and Space Physics, University of Colorado at Boulder, Boulder, Colorado 80309-0590, USA.
e-mail: peter.pilewskie@colorado.edu

1. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis. Fourth Assessment Report of the Intergovernmental Panel on Climate Change* http://ipcc-wg1.ucar.edu/wg1/Report/AR4WG1_Pub_Ch02.pdf (2007).
2. Ramanathan, V. *et al. Nature* **448,** 575–578 (2007).
3. Ramaswamy, V. *et al.* in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. *et al.*) 349–416 (Cambridge Univ. Press, 2001).
4. Holland, G. H. *et al. Bull. Am. Meteorol. Soc.* **82,** 889–901 (2001).
5. Chung, C. E., Ramanathan, V., Kim, D. & Podgorny, I. A. *J. Geophys. Res.* **110,** doi:10.1029/2005JD006356 (2005).
6. Mishchenko, M. I. *et al. Bull. Am. Meteorol. Soc.* **88,** 677–691 (2007).

MICROBIOLOGY

The inside story

Laurie E. Comstock

The human intestine is home to trillions of bacteria. Investigation of the colonization of the infant gut by these microorganisms is a prelude to understanding how they may act in both health and disease.

At birth, babies emerge from a sterile environment into one that is laden with microbes. The infant's intestine then rapidly becomes home to one of the densest populations of bacteria on Earth. Writing in *PLoS Biology*, Palmer *et al.*[1] report the most comprehensive analysis to date of the bacteria that first take up residence in the human intestine.

Interest in this ecosystem stems in part from the discovery of numerous benefits that arise from our intestinal microbiota: these bacteria help in extracting nutrients from food, and are instrumental in the development of the gut[2,3] and the immune system[4] after birth. However, gut microbes have also been linked to several disease states, including inflammatory bowel diseases and colon cancer, and less directly to maladies such as asthma, rheumatoid arthritis, atopic dermatitis and even autism[5,6]. An accurate and comprehensive analysis of the microbes present in the developing microbiota of the infant is an essential first step towards understanding which of them may affect the health of the host.

Palmer *et al.*[1] analysed the microbial composition of the intestinal ecosystem of 14 infants by sampling their faeces. Sampling began with the first stool after birth, and was followed by 25 further samples from each infant over their first year of life. The authors' method of quantifying the bacterial composition avoided the need to culture the bacteria. It involved use of a comprehensive DNA microarray that differentiated and quantified the distinct taxonomic groups present in the samples.

There are 22 broad taxonomic groupings, or phyla, of bacteria, but the bacteria abundant in the infant intestine fell into only three of them: the Gram-positive bacteria (Firmicutes and Actinobacteria), the Bacteroidetes and the Proteobacteria. Given the broad nature of these taxonomic groupings, the results are not entirely surprising — most of the bacteria known to associate with humans fall into these three major groupings. A previous analysis of the intestinal microbiota of healthy adults demonstrated the abundance of only two of these three phyla[7], with members of the Proteobacteria being only minor components. Proteobacteria are facultative anaerobes — that is, they can grow in the presence or absence of oxygen. They may be early settlers that are necessary to create the reduced environment required for the ensuing colonization by obligate anaerobes, which require oxygen-free conditions.

Contrasting with the similarity in the infants' microbiota at the phylum level, Palmer *et al.* found a remarkable degree of species-level variation, especially during the first few months. Some species appeared only transiently; others persisted for weeks to months. In general, there was no discernible pattern of abundant species or temporal mode of acquisition of particular organisms in different individuals. The two infants whose microbiotas were the most similar to each other were fraternal twins. These babies share both similar genetics and a similar environment. But their microbial profiles were no more like those of their own parents than they were to those of the parents of the other infants, implying that environment may play a greater role than genetics.

To try to identify the origin of the early colonizers, Palmer *et al.* investigated two relevant sources — the microbiotas of maternal vaginal fluid and of breast milk. The lactobacilli that predominate in the vaginal microbiota were not abundant in the early faecal samples, and the bacterial sample of only one infant 'clustered' with the vaginal microbiota of its mother, and then only during the first day after birth. So it seems that vaginal birth does not make a lasting contribution to an infant's intestinal microbiota. All babies in this study were breast-fed to some extent, but their intestinal microbiotas did not cluster with those of their mother's milk. Surprisingly, bifidobacterial species turned out to be only minor components: it is generally accepted that these bacteria are abundant in the stool samples of breast-fed infants, and that they are beneficial to their host.

Although the early microbiota of the gut seems largely to result from chance microbial encounters, by one year old there was a consistent convergence towards that of an adult-like microbiota, often coinciding with the introduction of solid food. Therefore, despite the absence of a programmed succession of early organisms, factors such as diet, gut development and environmental changes (possibly induced by the early colonizers) eventually result in the stable colonization of characteristic members of the adult microbiota. A notable exception was the lack of methanogenic archaea in the one-year samples. These organisms are abundant in the adult intestinal microbiota[7], where they consume methane produced by bacterial members.

Palmer and colleagues' methods[1] could be used to identify differences in the microbiotas of specific groups of infants and young children. A couple of examples suggest themselves.

First, an aspect only briefly mentioned in this paper is the effect of antibiotic treatment on microbial composition: although some antibiotics severely reduced the microbial load, the authors did not identify any consistent consequences of such treatment. Given the frequent use of antibiotics in children, this is a topic well worth following up with more focused and comprehensive analyses.

Second, it has been proposed that the increased prevalence of chronic inflammatory disorders in industrialized countries is associated with improved sanitation resulting in a reduced microbial burden, or with reduced exposure to particular microbes that regulate

the host's immune response[8]. The DNA microarray technique could be used to examine the microbiotas of infants in developing countries where the incidence of chronic inflammatory diseases is low compared with that in developed countries. The analysis of different cohorts of at-risk populations, followed over time, may eventually allow us to predict whether specific microbes increase the likelihood of a person developing a particular disease, or, conversely, whether the lack of a protective species leads to immune dysregulation. ■

Laurie E. Comstock is in the Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: lcomstock@rics.bwh.harvard.edu

1. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. *PLoS Biol.* **5,** e177 (2007).
2. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. *Science* **273,** 1380–1383 (1996).
3. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. *Proc. Natl Acad. Sci. USA* **99,** 15451–15455 (2002).
4. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. *Cell* **122,** 107–118 (2005).
5. von Mutius, E. *Immunobiology* **212,** 433–439 (2007).
6. Parracho, H. M., Bingham, M. O., Gibson, G. R. & McCartney, A. L. *J. Med. Microbiol.* **54,** 987–991 (2005).
7. Eckburg, P. B. *et al. Science* **308,** 1635–1638 (2005).
8. Strachan, D. P. *Br. Med. J.* **299,** 1259–1260 (1989).

APPLIED PHYSICS

Champing at the bit

Russell P. Cowburn

When two 'bits' of magnetic information race around a nanoscale wire, two factors determine whether or not they survive the course: the condition of the track, and how fast they respond to the starting signal.

Magnetic nanowires — wires of nanometre width made from magnetic materials — are an ideal playground for exploring the fundamentals of magnetism. Because they allow an easy flow of magnetic information, they could also be useful in computer chips: integrated circuits built from complex interconnected networks of nanowires have recently become the subject of intensive research. Writing in *Applied Physics Letters*[1], Hara *et al.* help to answer some crucial questions. What happens when nanowires carry high-density information at high speed? And how easy is it to synchronize the movement of many bits through the same nanowire without collisions and congestion?

The properties of virtually all magnetic materials are controlled by domains — extended regions where the spins of individual electrons are tightly locked together and point in the same direction. Where two domains meet, a domain wall forms. In 'soft' magnetic materials, these domain walls can be readily moved by applying a magnetic field. If such a material is formed into a narrow wire, a field can be used to push the walls from one end of the wire to the other, forming what amounts to a domain-wall conduit or conductor[2].

So far, this technique has been used to investigate the speed of the domain walls along the wire[3], as well as oscillations in the walls' internal structure[4] and the ability of electrical currents to move them[5]. By equating the direction of a domain's magnetization with a binary 0 or 1, a domain wall also becomes a mobile edge between data bits: the nanowire can thus be thought of as a physical means of transporting information in magnetic form.

This is an appealing development, because computers currently record information onto their hard disk in magnetic form. A new memory technology called magnetic random access memory (MRAM) that stores information in magnetic nanostructures has recently been launched. The convergence between electricity and magnetism in computing has thus already started — a trend known as spintronics[6]. Magnetic nanowires offer further convergence by allowing information to be transported and even processed[7] while in magnetic form.

But if these applications are to become an engineering reality, we must first understand how multiple domain walls within a nanowire network interact, especially when the walls are moving at high speed, and techniques need to be developed for the orderly procession of walls without collision. This is what Hara and colleagues[1] set out to do.

These authors effectively staged a race between two magnetic bits on a track formed from a circular nanowire racetrack. Using a static magnetic field, they set up a stable magnetic state in the ring comprising two domain walls on opposite sides of the ring, 180° apart. A rapid magnetic pulse applied to the ring causes both domain walls to move off at high speed in the same direction around the track. In an ideal world, both walls would start moving at the same instant and would move at the same speed. They should never catch up with each other.

In any real nanowire, however, there will be a variety of defects. These defects can cause variations in the threshold magnetic field at which a domain wall begins to move. Hara *et al.* vary the 'rise time' of the applied magnetic pulse — that is, how quickly the field strength grows — and thus investigate the conditions required for orderly transport of the domain walls (Fig. 1). Their argument is that if the rise time is slow, a domain wall that starts to move at a lower field strength gets more of a head start on its neighbour than it does if the rise time is fast. If the head start is comparable to half the ring's circumference, it would be possible for one domain

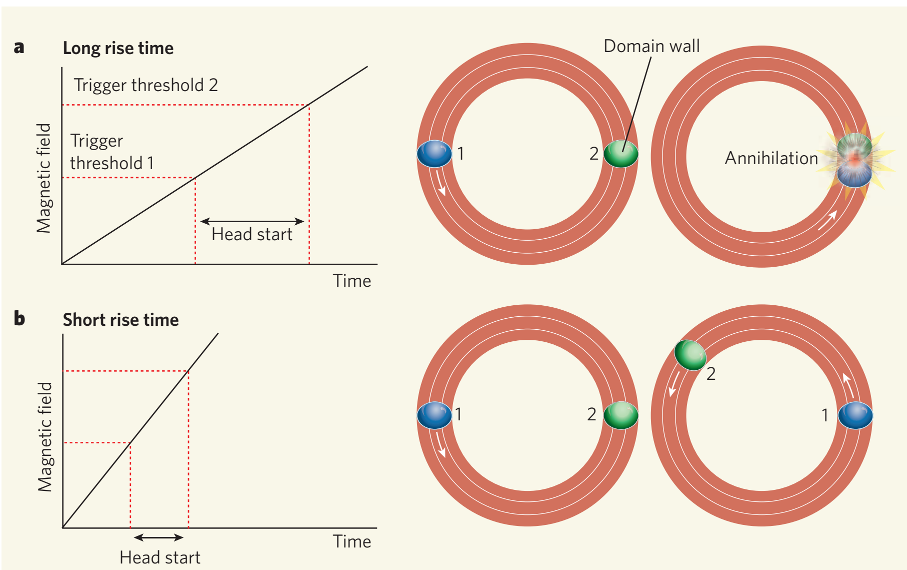


Figure 1 | At the nanoscale races. Hara *et al.*[1] use the rise time of a magnetic-field pulse to investigate the conditions under which magnetic domain walls (equivalent to 'bits' of magnetic information) move around a nanoscale track. In any realistic circuit, defects will cause different domain walls to begin to move at different threshold fields (here, the domain wall 1 moves at trigger threshold 1, and the domain wall 2 at trigger threshold 2). **a**, If the magnetic-field pulse rises only slowly (long rise time), the domain wall 1 is under starter's orders and away long before domain wall 2; if it manages half a circuit in this time, the two domain walls will collide and annihilate, resulting in the loss of magnetic information. **b**, With a quicker rise time, the head start of domain wall 1 is reduced, and domain wall 2 can pull away safely before domain wall 1 catches up.

wall to run into the back of the other.

The topology of the domain configuration is such that, in the event of a collision, the two walls would annihilate each other. Hara and colleagues' apparatus for measuring magnetic fields is sensitive enough to detect such annihilations. And their theory was indeed confirmed: for strong field pulses, annihilation occurs only if the rise time is slow. Moreover, a pulse with a rapid enough rise time is actually able to exchange the positions of the two walls, allowing each to travel a half lap of the ring without colliding.

The most important contribution of this work[1] is to highlight the way that the rise time of a driving stimulus and the distribution of defects in a nanowire interrelate in permitting, or hindering, the orderly transport of magnetic information. Hara *et al.* consider the way that naturally occurring defects can cause collisions between magnetic data; an interesting next step would be to see what synthetic defects do to the orderly flow of bits. A common trick used in magnetic nanowires is to decorate the edges with periodic notches to create a series of artifical interruptions to the free movement of a domain wall, with the aim of storing one data bit at each notch. The authors' approach could be useful for determining the risk of neighbouring bits annihilating when the signal is given for all domain walls to advance to the next notch. That might ultimately help those of us designing devices based on nanoscale networks to pack more data into already dense circuits. ■

Russell P. Cowburn is in the Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London SW7 2AZ, UK.
e-mail: r.cowburn@imperial.ac.uk

1. Hara, M., Kimura, T. & Otani, Y. *Appl. Phys. Lett.* **90,** 242504 (2007).
2. Ono, T. *et al. Science* **284,** 468–470 (1999).
3. Atkinson, D. *et al. Nature Mater.* **2,** 85–87 (2003).
4. Thomas, L. *et al. Nature* **443,** 197–200 (2006).
5. Lim, C. K. *et al. Appl. Phys. Lett.* **84,** 2820–2822 (2004).
6. Wolf, S. A. *et al. Science* **294,** 1488–1495 (2001).
7. Allwood, D. A. *et al. Science* **309,** 1688–1692 (2005).

Competing financial interests: declared (see online article for details).

CANCER

Broken genes in solid tumours

Matthew Meyerson

Mutations that cause portions of two genes to fuse together and form a hybrid gene are frequent in blood-related cancers. New findings implicate one such fusion gene in the most common type of lung cancer.

Lung cancer is the main cause of cancer deaths throughout the world[1], with an annual fatality of more than 1 million. Non-small-cell lung cancer (NSCLC) accounts for about 80% of all lung cancer cases. On page 561 of this issue, Soda *et al.*[2] report their discovery of a gene associated with human NSCLC. They find that a mutation involving positional rearrangement of genes along chromosome 2p activates the expression of the gene *ALK*, which encodes the ALK tyrosine kinase. Tyrosine kinases are molecular switches that regulate the activity of other proteins by adding phosphate groups to their tyrosine amino-acid residues. These enzymes have been implicated in many cancers, so blocking the activity of ALK kinase could lead to a powerful therapy for patients whose cancers bear this rearrangement.

Indeed, inhibiting protein products of activated oncogenes (cancer-associated genes) has been an effective strategy in cancer therapy[3], and many of the early successes have involved enzymatic blockade of tyrosine kinases. For example, the kinase inhibitor imatinib (Gleevec) prolongs the lifespan of patients with chronic myelogenous leukaemia[4], which is caused by chromosomal rearrangements that activate the ABL tyrosine kinase[5].

However, the strategy of gene-targeted therapy has had limited application, because so far only a fraction of the oncogenes that support the growth and survival of any given cancer type has been identified. Moreover, therapeutic strategies to disrupt gene function are not always available even for known oncogenes. Thus, Soda and colleagues' discovery of a new oncogene for a common cancer is momentous, especially given that they also found that the molecular structure of this oncogene's product makes it susceptible to enzymatic inhibition.

Soda *et al.*[2] obtained a lung-tissue specimen from a smoker NSCLC patient, isolated its total messenger RNA pool, and amplified it by generating a library of complementary DNA sequences to the mRNA transcripts. They then analysed the oncogenic activity of the complementary DNA sequences using a classic transformation assay. Through this assay, they searched for genes that could transform mouse 3T3 fibroblasts — that is, genes whose expression could confer on these cells certain properties of a tumour cell. Previously, several groups had attempted to use similar screens, but few obtained such dramatic results. So the successful use of this approach was a technical *tour de force*.

The authors identified a complementary DNA sequence derived from a fusion mRNA transcript as the sequence that could transform 3T3 cells. The first part of this sequence consisted of a portion of the *EML4* gene, the second of portions of the *ALK* gene. Originally, *ALK* was identified through its role in anaplastic large-cell lymphoma, a type of blood-derived cancer[6]; the discovery of its activation in lung cancer comes as a surprise[2].

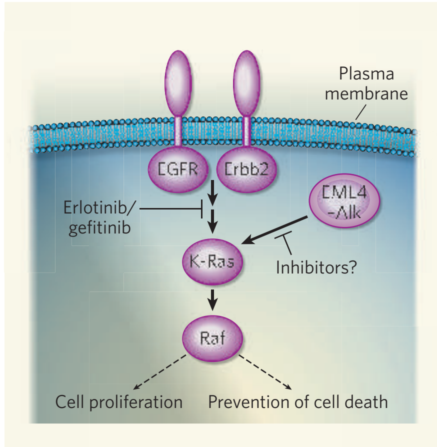


Figure 1 | The EML4–ALK fusion protein and lung cancer. Mutations in two tyrosine kinases — EGFR and Erbb2 — as well as in signalling molecules downstream of these receptors (Ras and Raf), have been implicated in non-small-cell lung cancer (NSCLC). The mutant proteins are thought to stimulate the proliferation of cancer cells and/or prevent their programmed death. Soda *et al.*[2] find that in some cases of NSCLC, a chromosomal rearrangement results in a fusion gene, the product of which is EML4–ALK. This fusion protein also functions as an activated tyrosine kinase, and thus might stimulate the EGFR-mediated signalling pathway. Tyrosine kinase inhibitors such as erlotinib and gefitinib are effective treatments for NSCLC, and inhibitors of EML4–ALK might therefore be equally promising drugs in the therapy of this cancer.

Most NSCLCs, especially lung adenocarcinomas — the commonest variety according to examinations of tissue specimens — involve the activation of a molecular signalling pathway mediated by receptor tyrosine kinases such as EGFR and Erbb2 and their downstream Ras and Raf oncogenes[7]. So it is plausible that activated ALK, like EGFR/Erbb2, might constitutively switch on the Ras–Raf signalling pathway (Fig. 1).

Soda and colleagues then searched for the incidence of the *EML4–ALK* fusion gene in Japanese NSCLC patients; of the 75 patients studied, 5 carried this oncogene. The incidence of *ALK* rearrangements in other patient populations remains an open and important question, because the frequencies of two other mutations — in *EGFR* and *K-RAS* — that also lead to NSCLC vary between patients of different ethnicity. Among east Asian NSCLC patients, about 40% harbour *EGFR* mutations and 10% carry *K-RAS* mutations; however, in Europe and North America only about 10% of NSCLC patients carry *EGFR* mutations, and in about 25% the *K-RAS* oncogene is mutated.

50 & 100 YEARS AGO

50 YEARS AGO

Physics and Applied Mathematics — The contributions to this excellent survey were originally given as lectures at a symposium arranged in 1954... where a number of the leading physicists in the United States explained the recent advances in their own fields... Thus, H. A. Bethe mentions the plans then being made by Reines and Cowan to detect the neutrino by inverse beta-decay; C. H. Townes explains the principle of the 'maser' (as yet unnamed), being rather guarded as to whether it will eventually work usefully... W. Shockley gives a fairly elementary but full description of transistor physics... I. I. Rabi discusses atomic structure, emphasizing the importance of recent work on positronium... V. F. Weisskopf analyses the complicated pattern of the elementary particles, and forecasts more to come as cosmotron work in the ultra-high energy range continues... E. U. Condon, on physics and the engineer... [directs] attention to the shortage of high-school science teachers. This last appears to be a feature of the times, and not restricted to any one country; one wonders if it really is the only problem that is quite unapproachable by the methods of operational research, for there must be some way of averting the impending academic dust-bowl.

From *Nature* 3 August 1957.

100 YEARS AGO

The experiments which have been in progress in the Congo State for some time past in training the African elephant for domestic work are progressing satisfactorily. During the first three months of the present year, eight young elephants were captured, bringing the stud up to a total of thirty. At first the wild elephants suffer in health on confinement, but this depression soon passes off. Similar experiments are being made in the British territory of Uganda, but so far the results there are uncertain.

From *Nature* 1 August 1907.

In the small sample set studied by Soda *et al.*, mutations in *ALK* were mutually exclusive with *K-RAS* and *EGFR* mutations.

So far, the tyrosine kinase inhibitors gefitinib (Iressa) and erlotinib (Tarceva) are the only effective targeted therapies for lung cancer and are approved in different countries; these drugs seem to be most effective for the treatment of NSCLC patients with *EGFR* mutations[8–10]. Soda and colleagues present evidence that ALK inhibition is physiologically feasible, showing that mouse BA/F3 cells that were transformed with the *EML4–ALK* fusion gene can be specifically killed with an ALK inhibitor. The potential therapeutic use of such inhibitors would be of particular interest if the *EML4–ALK* fusions are often found in smoker NSCLC patients; this is because gefitinib/erlotinib are most effective for non-smoker NSCLC patients whose tumours are most probably due to *EGFR* mutations.

Until recently, chromosomal rearrangements have mainly been linked to blood-related cancers and seldom to solid tumours. However, the discovery of *TMPRSS2–ERG* and *TMPRSS2–ETV1* fusion genes in prostate cancer[11,12], and now the implication that the *EML4–ALK* fusion gene is involved in lung cancer, indicate that activated fusion genes associated with chromosomal rearrangements are probably both common and important in solid tumours. Developing systematic genome-wide approaches for discovery and diagnosis using these gene fusions will probably provide a leap forward in our understanding of the causes of solid-tumour cancers. ■

Matthew Meyerson is at the Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA.
e-mail: matthew_meyerson@dfci.harvard.edu

1. Parkin, D. M., Bray, F., Ferlay, J. & Pisani P. *CA Cancer J. Clin.* **55,** 74–108 (2005).
2. Soda, M. *et al. Nature* **448,** 561–566 (2007).
3. Sawyers, C. *Nature* **432,** 294–297 (2004).
4. Druker, B. J. *et al. N. Engl. J. Med.* **344,** 1031–1037 (2001).
5. de Klein, A. *et al. Nature* **300,** 765–767 (1982).
6. Morris, S. W. *et al. Science* **263,** 1281–1284 (1994).
7. Sharma, S. V., Bell, D. W., Settleman, J. & Haber D. A. *Nature Rev. Cancer* **7,** 169–181 (2007).
8. Lynch, T. J. *et al. N. Engl. J. Med.* **350,** 2129–2139 (2004).
9. Paez, J. G. *et al. Science* **304,** 1497–1500 (2004).
10. Pao, W. *et al. Proc. Natl Acad. Sci. USA* **101,** 13306–13311 (2004).
11. Tomlins, S. A. *et al. Science* **310,** 644–648 (2005).
12. Tomlins, S. A. *et al. Nature* **448,** 595–599 (2007).

MAGNETISM

Managed mess

Zachary Fisk

The presence of non-magnetic atoms can create a random internal field in magnetic crystals. Tuning that field from outside allows the intrinsic magnetic properties of the material to be precisely controlled.

The eccentric inventor and industrialist Diet Smith — financier of Dick Tracy's crime-fighting gadgetry (Fig. 1) — noted long ago that "he who controls magnetism controls the Universe". And efforts to control magnetism continue to make progress, as Silevitch *et al.* ably demonstrate in this issue[1] (page 567). These authors use a combination of chemical impurities and an external, variable magnetic field to manipulate the innate magnetism of a crystalline ferromagnet.

In a material with magnetic ordering, the directions of the magnetic moments that reside on every atom have a regular pattern. Among such materials, one expects ferromagnets (the atomic moments of which are all oriented in the same direction, even in the absence of an external magnetic field) to be simpler than antiferromagnets (which have some or all adjacent moments pointing in non-parallel directions). But being simpler is not the same as being simple. The physics of some ferromagnets is simpler than that of others, and the genius of he who wishes to control magnetism lies in finding these materials.

Silevitch and colleagues' choice[1] is the insulating ferromagnet $LiHoF_4$ (ref. 2). Magnetically, the $LiHoF_4$ crystal can be thought of as a three-dimensional stack of planar sheets of bar magnets made of holmium (Ho) atoms, which carry an intrinsic atomic magnetic moment. Each of these bar magnets can be oriented only with its north pole pointing at right angles — either up or down — to the plane of the sheets. This arrangement is known as an Ising system. Below this material's critical Curie temperature for ferromagnetic ordering — 1.53 kelvin — all of the north poles point in the same direction. The simplicity of this material is due partly to the fact that the dominant magnetic coupling between the atomic magnets is exactly the same as the dipole–dipole interaction of bar ferromagnets seen in basic school experiments.

The authors disturb this perfect simplicity by 'diluting' their crystal, replacing some of the holmium atoms in the crystal lattice with atoms of the non-magnetic, but chemically almost equivalent, element yttrium (Y), so that only a fraction x of the original holmium is left. The resulting material has the chemical formula $LiHo_xY_{1-x}F_4$. This breaking of the symmetry

TRIBUNE MEDIA SERVICES, INC.

Figure 1 | Speaking clock. **Dick Tracy's two-way wrist radio was the most obvious consummation of his backer's faith in the power of magnetism.**

surrounding each holmium site leads to 'off-diagonal' coupling of spin components between them, which results in a spatially random internal magnetic field[3]. At the heart of Silevitch and colleagues' work lies the insight that they can control the amplitude of this random field — and so use it as an experimental variable — by applying an external magnetic field in the plane of the holmium bar magnets.

The system thus created provides a playground in which to explore the changing behaviour of a magnetic material near critical points in response to variations in parameters such as temperature, magnetic field and pressure (this paper[1] concentrates on the first two of these variables). There are two different critical points of interest: the classical ferromagnetic Curie temperature, which occurs when the external field is zero; and the quantum critical point, which occurs when the Curie temperature has been driven to absolute zero by varying an experimental parameter such as the external magnetic field or the chemical doping.

The authors show that the immediate effect of yttrium substitution in $LiHoF_4$ is to depress the ferromagnetic Curie temperature linearly with decreasing holmium proportion, down to a proportion $x = 0.2$. Of particular interest is the effect of an applied magnetic field H on the material's magnetization M. This is defined by a parameter known as the magnetic susceptibility, $\chi = \mathrm{d}M/\mathrm{d}H$ (the rate of change of magnetization in a small applied magnetic field). In these experiments, χ also depends on the transverse applied field used to tune the random fields at the holmium sites.

Silevitch *et al.*[1] extract inverse exponential power-law dependencies for χ both as a function of temperature (at zero effective transverse magnetic field) and as a function of the effective transverse magnetic field (at the Curie temperature). A comparison of the pure compound ($x = 1$) with a diluted variant ($x = 0.44$) at zero applied magnetic field finds that the first power law is essentially unaltered, but that the second is considerably different in the diluted material. Remarkably, the authors find that this difference, which is the effect of random fields, can be parametrized in a modified Curie law, the standard expression that relates a material's magnetization to temperature and applied magnetic field. All that is needed are extra terms that take into account the suppression of magnetic order through chemical dilution and transverse fields, as well as the effects of quantum fluctuations and higher-order random field effects.

The immediate import of this work is twofold. First, the authors have shown how random field effects in a ferromagnet can be experimentally seen and tuned in a simple and standard laboratory measurement. Second, they have, for the first time, found a quantitative route into the realm of Griffiths singularities[4]. This term describes anomalies expected to occur in the magnetization (as a function of temperature at zero field) of randomly diluted Ising ferromagnets below the Curie temperature of the undiluted material. The extensive theoretical literature on the subject suggests that the difficulty in observing such singularities is caused by their being extremely weak, because they arise from isolated still-pure regions of the diluted system that occur with very low probability.

But these results[1] are also a prod to thinking about where else, and in what other related classes of material, such tuning of magnetic properties might be possible. The results could have relevance, for example, to the control of the magnetic domain boundaries, a focus of enormous research interest (see, for example, page 544). He who controls magnetism might not end up controlling the world, but will have a better handle on technologies so important in our information age. ■

Zachary Fisk is in the Department of Physics and Astronomy, University of California Irvine, 2186 Frederick Reines Hall, Irvine, California 92697, USA.
e-mail: zfisk@uci.edu

1. Silevitch, D. M. *et al. Nature* **448,** 567–570 (2007).
2. Cooke, A. H., Jones, D. A., Silva, J. F. A. & Wells, M. R. *J. Phys. C* **8,** 4083–4088 (1975).
3. Tabei, S. M. A., Gingras, M. J. P., Kao, Y.-J., Stasiak, P. & Fortin, J.-Y. *Phys. Rev. Lett.* **97,** 237203 (2006).
4. Griffiths, R. B. *Phys. Rev. Lett.* **23,** 17–19 (1969).

CANCER

An infernal triangle

Alberto Mantovani

Signals induced by sex hormones and inflammation have been viewed as different aspects of tumour development. But a three-way interaction between these two classes of signal and carcinogenesis has emerged.

In 1896, George Beatson discovered that surgical removal of the ovaries slows the growth of breast tumours, so revealing a link between sex steroid hormones and cancer of their target endocrine organs. Moreover, it has long been known that females are less susceptible to tumours at sites that are not conventional target organs of sex steroid hormones, such as the gastrointestinal tract[1]. Differences in drinking and smoking habits do not seem to fully account for these gender differences. Now, several studies open up unexpected vistas on the relationship between gender differences and cancer. They suggest that the inflammatory response, which is mediated by the innate, or nonspecific, immune system, might be an essential element of the action of sex steroid hormones.

Hepatocellular carcinoma, the most common type of liver cancer, is a frequent outcome of chronic inflammation triggered by hepatitis due to viral infection[2], with males being more susceptible to this tumour than females. Naugler *et al.*[3] have now explored the mechanisms underlying this gender difference. They started from the observation that, in response to injections of the carcinogen diethylnitrosamine, male mice had higher levels of a cell signalling compound (cytokine) known as interleukin-6 (IL-6), which is a marker as well as a mediator of chronic inflammation. The sex difference in susceptibility to cancer was absent in mice deficient in IL-6 and in the adaptor protein MyD88. The latter operates downstream of the Toll-like receptors that sense microbial invasion and tissue damage, or (in the form of the IL-1 receptor) act as amplifiers of inflammation.

The following picture emerges from *in vitro* work and from *in vivo* studies in mice (Fig. 1a). Carcinogen-induced tissue damage results in the release of debris, which in turn causes MyD88-dependent activation of Kupffer cells in the liver. These cells are a type of macrophage, a main player in the innate immune response. The cellular debris probably acts via Toll-like receptors, although this has not been formally proven. The Kupffer cells produce IL-6, which promotes liver injury, inflammation, compensatory cell proliferation, and carcinogenesis. In females, however, oestrogen steroid hormones act through gene transcription factors (such as NF-κB) to inhibit IL-6 production in Kupffer cells, and so to protect female mice from cancer. Curiously, tumour-necrosis factor, another cytokine, has been implicated in liver carcinogenesis[2,4,5] but is not involved in this gender difference.

Other investigations offer further evidence on the role of innate immune signalling in carcinogenesis. Mice deficient in MyD88 are relatively resistant to intestinal cancers, and

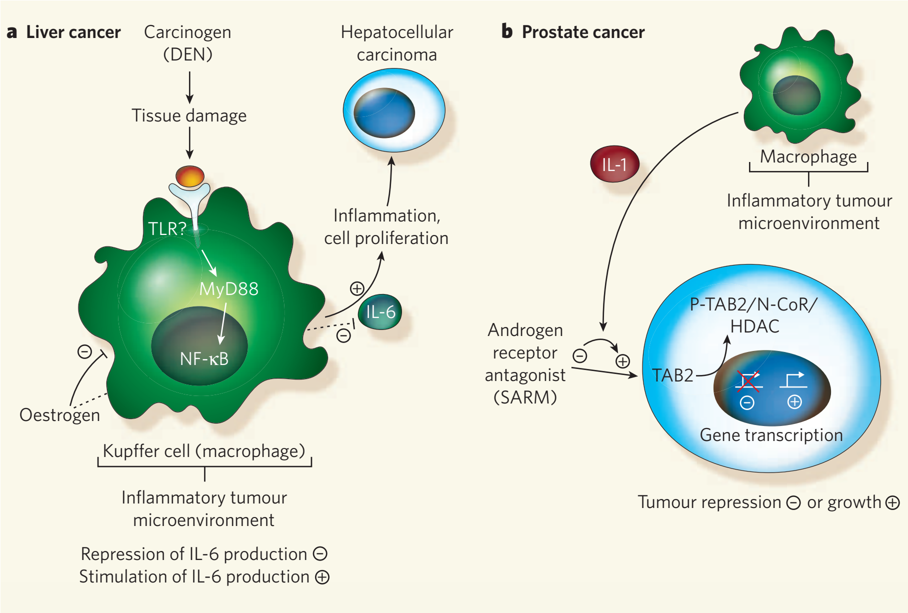


Figure 1 | Sex steroid hormones, inflammation and cancer. a, Naugler *et al.*[3] show that, in the liver, a carcinogen (diethylnitrosamine, DEN) causes tissue damage that (presumably acting via Toll-like receptors, TLR) activates the MyD88–NF-κB signalling pathway in Kupffer cells. These cells, a type of macrophage, produce interleukin-6 (IL-6), which in turn promotes inflammation, tissue damage, cell proliferation and tumour formation. Oestrogens interfere with NF-κB activity and IL-6 production, so females tend to be protected against liver cancer. **b,** Zhu *et al.*[10] find that, in prostate cancer, macrophage-derived IL-1 converts an androgen-receptor antagonist (a steroid androgen receptor modulator, or SARM) into an agonist that stimulates gene transcription. The molecular sensor for the IL-1 inflammatory signals is TAB2, which on phosphorylation releases the TAB2/N-CoR/HDAC repressor complex from the gene promoter sequence, activating gene transcription. A SARM is thus converted from a tumour inhibitor to a tumour promoter.

this resistance is associated with altered expression of NF-κB-dependent modifier genes[6]. In complementary studies[7,8], mice deficient in an inhibitor of Toll-like-receptor signalling, called TIR8 or SIGIRR, showed increased susceptibility to cancer associated with colitis (inflammation of the colon). Thus, oestrogens interfere with carcinogenesis by acting on NF-κB, an endogenous tumour promoter[9] that is activated by MyD88-dependent signalling in the innate immune system.

Zhu *et al.*[10] have provided an unrelated line of evidence for the interplay between sex steroid hormones and inflammation in prostate cancer, a classic hormone-stimulated tumour. They looked at the actions of macrophages in prostate cancer cells *in vitro*, in the context of signalling through the androgen receptors in these cells. Androgens are male sex hormones, the activity of which is modulated by agents called 'selective androgen-receptor modulators' (SARMs), which act as weak agonists or antagonists (so amplifying or blocking hormone action). Their counterparts for oestrogen are 'selective estrogen-receptor modulators' (SERMs). Each type of molecule is used in different circumstances to stimulate or repress their respective receptor action, and so control the action of oestrogens and androgens.

Working with cells *in vitro*, Zhu *et al.* found that the macrophages in prostate cancer produce IL-1, which converts SARMs from their intended function of inhibitors of androgen-receptor-induced gene expression to activators of expression (Fig. 1b). The mechanism involves a protein, TAB2, which acts as a sensor for inflammatory signals. TAB2 is a component of a repressor complex known as TAB2/N-CoR/HDAC: inflammation results in the phosphorylation of TAB2, lifting repression and unleashing gene transcription. Thus, an inflammatory microenvironment can convert a SARM from a Dr Jekyll into a Mr Hyde.

The studies discussed here provide firm evidence for an unexpected feature of the connection between inflammation and cancer[2,5,11] — a bidirectional interaction between inflammatory mediators and sex steroid hormones. Macrophages are components of the inflammatory microenvironment of tumours[5], and are apparently central players in the link between inflammation and sex hormones in carcinogenesis. Such a link may reflect the interplay between inflammation and hormones in reproduction.

For the next steps in this research, it will be necessary to find out whether these mechanisms operate not just *in vitro* and in mice, but in human tumours as well. SARMs, and SERMs such as tamoxifen, which is a treatment for breast cancer, are widely used in the clinic. The new results call for an analysis of their mode of action, and mechanisms of resistance to them, that take account of the inflammation connection. Such studies may pave the way to strategies for targeting the liaisons between inflammation and sex steroid hormones for therapeutic ends. ■

Alberto Mantovani is at the Istituto Clinico Humanitas, Via Manzoni 56, 20089 Rozzano, Milan, and the Istituto di Patologia Generale, Università degli Studi di Milano, Milan, Italy.
e-mail: alberto.mantovani@humanitas.it

1. Levi, F., La Vecchia, C., Lucchini, F. & Negri, E. *World Health Stat. Q.* **45,** 117–164 (1992).
2. Balkwill, F. & Mantovani, A. *Lancet* **357,** 539–545 (2001).
3. Naugler, W. E. *et al. Science* **317,** 121–124 (2007).
4. Pikarsky, E. *et al. Nature* **431,** 461–466 (2004).
5. Balkwill, F., Charles, K. A. & Mantovani, A. *Cancer Cell* **7,** 211–217 (2005).
6. Rakoff-Nahoum, S. & Medzhitov, R. *Science* **317,** 124–127 (2007).
7. Xiao, H. *et al. Immunity* **26,** 461–475 (2007).
8. Garlanda, C. *et al. Cancer Res.* **67,** 6017–6021 (2007).
9. Karin, M. *Nature* **441,** 431–436 (2006).
10. Zhu, P. *et al. Cell* **124,** 615–629 (2006).
11. Coussens, L. M. & Werb, Z. *Nature* **420,** 860–867 (2002).

GENOMIC BIOLOGY

The epigenomic era opens

Stephen B. Baylin and Kornel E. Schuebel

Readout of information from the genome depends on intricate regulation of how DNA is packaged by proteins. The great endeavour to reveal how this packaging operates pan-genomically is now under way.

A new era is opening for biologists involved in understanding cellular systems. It is exemplified by papers by Mikkelsen *et al.* (page 553 of this issue)[1] and Barski *et al.* (published in *Cell*)[2] — they describe the kind of unprecedented insights that are emerging from investigations of how a single mammalian genome can be regulated to produce different cell types.

The technical and biological advances described in these studies extend the remarkable accomplishments of elucidating the structure[3], then the sequence[4,5], of the human genome; and they reflect a growing, 'post-genomic', appreciation of the complexities of genome structure and function (Fig. 1). The intriguing — and daunting — challenge now is to understand the process of how and when specific DNA regions are controlled to produce the cellular diversity that underpins the development and maintenance of a single organism.

Central to this challenge is the task of enumerating the dizzying number of proteins interacting with the genome, and the functions they subserve. These proteins, called histones, form a combination with DNA that is termed chromatin. It is chromatin that provides the software packaging for the readout of the DNA hard drive. If alterations in genome heritable states occur through a change in the hard drive (that is, through a change in the primary sequence of DNA), a genetic alteration or mutation has occurred. This contrasts with an epigenetic change, which is an alteration in the heritable states of DNA function produced by altering the chromatin software. Epigenetic changes lie at the heart of how organisms generate different types of tissue under different circumstances — in embryonic development, in regulating cell renewal in adults, and in the cellular responses of the organism to environmental factors and stress. Moreover, disease states such as cancer are associated with a combination of both genetic and epigenetic abnormalities.

The central unit of chromatin is the nucleosome, which is constructed from short regions of DNA wound around an octet of histone proteins. This unit can modulate the readout from DNA in at least three ways.

First, nucleosomes can be physically rearranged on DNA by complexes known as chromatin-remodelling proteins[6] — generally, the greater the distance between nucleosomes, and so the 'openness' of chromatin, the higher the likelihood that such regions of DNA will be transcribed into RNA. Second, many nucleosomes can be compacted into higher-order aggregates to form 'closed' chromatin, or heterochromatin[6], thereby preventing transcription. The balance between the open and closed parts of the genome facilitates proper gene-expression patterns in given cell types, and also prevents unwanted gene transcription.

Third, there is a complex interplay between enzymes that can modify particular amino acids in the histone component of the nucleosomes, and those that reverse the modifications. The modifications, or histone 'marks', interact with proteins that bind to and interpret them. The marks were initially seen as a 'histone code', the idea being that a restricted number of them would specify the 'on' or 'off' state of RNA production from DNA[7]. This concept was a most useful starting point. But it is increasingly recognized that the constituents of chromatin, and nucleosome structure, position and modification, are highly complex. It is a balance between these factors that marks an individual gene, or groups of genes, for various levels and states of expression[8]. That is, there is no simple on–off code.

All of which brings us to the papers by Mikkelsen *et al.*[1] and Barski *et al.*[2]. Both represent examples of genome 'tiling' approaches — the aim being to catalogue, across the entire human genome, the locations not only of key histone modifications but also of proteins that respond to and mediate them. Mikkelsen *et al.* begin the process of mapping how these parameters change as cells negotiate their conversion from immature to adult states, whereas Barski *et al.* examine a more mature cell state. The two groups used an ingenious new technology, Solexa 1G sequencing, which allows millions of short DNA 'sequence tags' to be assigned to individual histone marks, thus mapping the marks to their precise location in the genome.

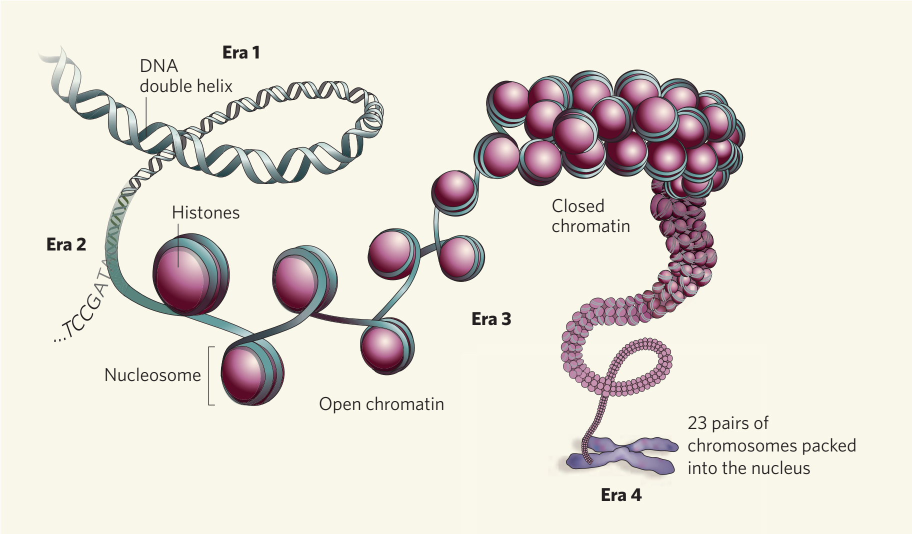


Figure 1 | Genomic architecture, and eras of investigation. Eras 1 and 2 were launched by the elucidation of the structure of DNA, and then of the sequence of the human genome. DNA is packaged with histone proteins to form chromatin, the central unit of which is the nucleosome. As examples of work in the unfolding era 3 — the era of the epigenome — the papers of Mikkelsen *et al.*[1] and Barski *et al.*[2] provide genome-wide linear maps relating histone modifications ('marks') to the active or inactive sites where various types of regulatory RNA, as well as messenger RNA, are produced. Nucleosomes are further organized to create open and closed regions of chromatin, which in turn create three-dimensional structures that encompass different levels of gene organization in the chromosomes packed into a cell nucleus. A likely era 4 will involve generation of maps of this 'genome topography' to reveal how a single genome can produce such a diversity of cell types.

The results are remarkably comprehensive linear maps of the principal chromatin constituents across the human genome. The maps highlight the complexity of DNA packaging, and reveal that combinations of histone modifications and positions, rather than single histone marks, correlate most accurately with multiple levels of the genes' transcriptional states. Histone characteristics can define the immediate start sites of genes, which are often regulatory in nature. But they can also define discrete but distant regions that influence gene expression, as well as regions that may encompass an entire gene to prompt its active or repressed transcription.

The papers also provide insights about genomic regions — within genes, or between genes — that are unexpectedly marked for expression activity. These data relate to the recent revelations that much more of our genome than previously thought is engaged in expressing RNA from DNA. The result is production not only of 'classical' messenger RNAs (which produce the proteins defined by the initial analyses of the genome sequence), but also of a huge number of regulatory RNAs (which modulate genome readout by producing multiple forms of the same protein or without producing proteins at all[9]).

So, are we done with mapping the genome? Hardly. These genome-packaging data[1,2] provide a first linear view that can only hint at the three-dimensional aspects of how the genome is organized in the cell nucleus to regulate DNA. We already know, broadly at least, that nucleosome-mediated chromatin domains create three-dimensional structures that surround individual gene-regulatory regions, whole genes, groups of genes and genes encompassed in chromosome territories[10] — producing, altogether, what can be seen as a genome topography. Perhaps a complete view of the genome will require a further era of investigation (Fig. 1), in which we generate maps of the genomic topography that characterizes each of the many cell types of which we are constituted. Who knows? We are just at the beginning of exploring how a single genome can spawn multiple epigenomes. ■

Stephen B. Baylin and Kornel E. Schuebel are at the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, 401 North Broadway, Baltimore, Maryland 21231, USA.
e-mail: sbaylin@jhmi.edu

1. Mikkelsen, T. S. *et al. Nature* **448,** 553–560 (2007).
2. Barski, A. *et al. Cell* **129,** 823–837 (2007).
3. Watson, J. D. & Crick, F. H. C. *Nature* **171,** 737–738 (1953).
4. Lander, E. S. *et al. Nature* **409,** 860–921 (2001).
5. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
6. Li, B., Carey, M. & Workman, J. L. *Cell* **128,** 707–719 (2007).
7. Jenuwein, T. & Allis, C. D. *Science* **293,** 1074–1080 (2001).
8. Bernstein, B. E., Meissner, A. & Lander, E. S. *Cell* **128,** 669–681 (2007).
9. Kapranov, P. *et al. Science* **316,** 1484–1488 (2007).
10. Albiez, H. *et al. Chromosome Res.* **14,** 707–733 (2006).

Q&A

PANORAMIC IMAGES/GETTY

Alpine ecosystem. Species in mountain habitats are especially sensitive to climate change.

BIODIVERSITY

Climate change and the ecologist

Wilfried Thuiller

The evidence for rapid climate change now seems overwhelming. Global temperatures are predicted to rise by up to 4 °C by 2100, with associated alterations in precipitation patterns. Assessing the consequences for biodiversity, and how they might be mitigated, is a Grand Challenge in ecology.

How serious is climate change compared with other factors affecting biodiversity?

Very — but it tends to act over a longer timescale. The ecological disruption wrought by climate change is generally slower than that caused by other factors. Such factors include habitat destruction through changes in land use; pollution, for example by nitrogen deposition; the invasion of ecosystems by non-native plant and animal species (biotic exchange); and the biological consequences of increased levels of carbon dioxide in the atmosphere (Fig. 1, overleaf). In the short-to-medium term, human-induced fragmentation of natural habitat and invasive species are particular threats to biodiversity. But looking 50 years into the future and beyond, the effects of climate are likely to become increasingly prominent relative to the other factors.

What are the effects of climate change?

Most immediately, the effects are shifts in species' geographical range, prompted by shifts in the normal patterns of temperatures and humidity that generally delimit species boundaries. Each 1 °C of temperature change moves ecological zones on Earth by about 160 km — so, for example, if the climate warms by 4 °C over the next century, species in the Northern Hemisphere may have to move northward by some 500 km (or 500 m higher in altitude) to find a suitable climatic regime. Higher temperatures are likely to be accompanied by more humid, wetter conditions, but the geographical and seasonal distribution of precipitation will change. Summer soil moisture will be reduced in many regions such as the Mediterranean basin, thus increasing drought stress. Overall, the ability of species to respond to climate change will largely depend on their ability to 'track' shifting climate through colonizing new territory, or to modify their physiology and seasonal behaviour (such as period of flowering or mating) to adapt to the changed conditions where they are.

What about the effect of atmospheric gases?

Carbon dioxide is, of course, known as one of the main drivers of the greenhouse effect, and so of increasing temperatures. But it is also essential for green-plant photosynthesis. Increased atmospheric CO_2 results in an increase in photosynthesis rates (through CO_2 fertilization), which could potentially balance the effect of temperature increase. This has the largest effect in regions where plant growth is limited by the availability of water, and will probably alter the competitive balance between species that differ in rooting depth, photosynthetic pathway or 'woodiness', as well as the subterranean organisms associated with them. Likewise, an increase of anthropogenic atmospheric nitrogen deposition affects nitrogen-limited regions (temperate and boreal forests, and alpine and Arctic regions) by conferring a competitive edge on plants with high maximum growth rates.

Which ecosystems are we talking about?

All of them, but climate change will affect them in different ways. For example, in marine ecosystems the possible consequences include increased thermal stratification (in which temperature differences separate water layers), reduced upwelling of nutrients, decreased pH and loss of sea ice. These changes will influence the timing and extent of the spring bloom of phytoplankton, and so the associated food chain (krill to fish to marine mammals and birds). On the terrestrial side, deserts, grasslands and savannahs in temperate regions are likely to respond to changes in precipitation and warming in various ways. Mediterranean-type ecosystems, which occur worldwide and are characterized by shrublands, are especially sensitive, as increased temperature and drought favour development of desert and grassland. In tropical regions, CO_2 fertilization — in which plants absorb carbon from the atmosphere — and altered patterns

of naturally occurring fires will have a strong influence. On tundra, low-growing plants are especially important as habitats for other organisms: their poleward movement will have an ecosystem-wide impact. Finally, species living on mountains are particularly sensitive to changed conditions, in that migration upwards can occur to only a limited extent.

How do biologists monitor changes in biodiversity?

Long-term observations and re-surveys of previously sampled sites are traditional approaches. In certain areas, natural-history societies have long recorded the seasonal time of appearance (of flowers, for instance, or migratory birds), or species' ranges. Such data sets are then viewed against measured variations in temperature or precipitation. Another approach is the re-survey of sites sampled 50 or 100 years previously. Species' identities and abundances are then compared with changes in such external factors as climate or land use. The drawback of both approaches lies in distinguishing a true cause from a correlation.

Do experimental studies help?

Monitoring programmes can be complemented by research in microcosms or, for example, on existing plots of grassland or forest. In these experiments, temperature, precipitation and even CO_2 concentration can be manipulated, and such work often reveals unexpected responses arising from the complex interplay of different factors. But for obvious reasons these experiments are difficult to carry out on large spatial and temporal scales.

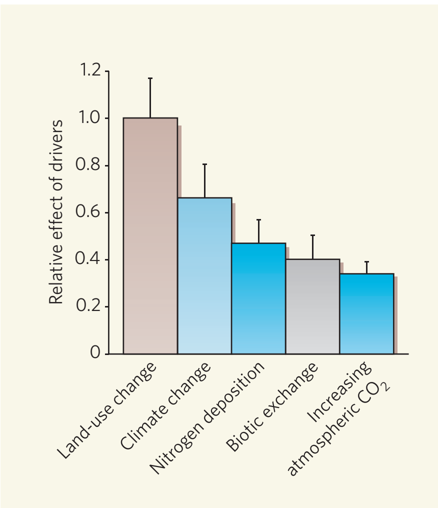


Figure 1 | The main factors, or 'drivers', affecting biodiversity. This summary of the relative effects by the year 2100 is a composite derived from calculations carried out for 12 individual terrestrial and freshwater ecosystems by O. E. Sala *et al.* (*Science* 287, 1770–1774; 2000). Overall, changes in land use constitute the main estimated impact on biodiversity, but the pattern varies considerably for different ecosystems. According to Sala and colleagues' calculations, climate change will have the strongest effect on Arctic, alpine and boreal ecosystems, whereas biotic exchange (that is, invasion by non-native species) will exert its main influence in lakes.

What responses to climate change are actually documented?

In the Northern Hemisphere, the range of terrestrial plants and animals has shifted, on average, 6.1 km per decade northward or 6.1 m per decade upwards, with advance of seasonal phenomena by 2.3–5.1 days per decade over the past 50 years. These changes are significantly correlated with measured changes in temperature and precipitation. The relationships are correlative in essence, but are too robust, numerous and consistent to be random or to have arisen from other factors (such as natural climatic variability or land-use change). Similarly, the remarkable increase in the plant diversity of some high-elevation peaks in Switzerland over the past 100 years, owing to the upward shift of species that traditionally inhabited lower elevations, can be attributed to changed climate regimes.

Is there a consistent global picture?

We can only guess that patterns such as these are likely to be global in compass, but to differing extents. The two poles are probably being most affected, because the greatest changes in temperature and precipitation are occurring there. By contrast, biodiversity in the equatorial belt is likely to suffer more immediately from deforestation and land degradation. Most of the detailed quantitative studies come from the Northern Hemisphere, or from well-studied 'hotspots' of biodiversity such as the Cape Floristic Region in South Africa. Even in these regions, it is difficult to disentangle the effects of climate change from those of other factors. And we have little or no data on vast swaths of territory in South America, Africa and Asia.

Do climate change and other factors interact?

They do. A notable example concerns invasive species: change in climate can trigger change in biodiversity by creating opportunities for previously innocuous alien species by enhancing their reproductive capacity, their survival and their competitive power against the native flora and fauna. The dispersal of many species, including microorganisms, has been immeasurably increased by the globalization of human economic activity and trade. A combination of climate change, species invasions and reduced areas of natural habitat is likely to promote biotic homogenization in biodiversity hotspots in particular, and to foster unpredictable interactions between plants, animals and microorganisms.

How do ecologists set about forecasting the impacts of climate change on biodiversity?

Experimental studies are informative, but can rarely be generalized. Another approach is to combine ecological modelling with various scenarios of climate change. For example, statistical 'niche-based' models are used to determine the environmental conditions that currently account for species' distributions, and the results can be compared with models of future climate and patterns of land use to predict where these conditions will occur in the future. Validations are usually done by modelling past distributions (as, for instance, surmised for plants from a pollen database). These models don't take into account biological factors such as competition and evolutionary history, but have produced forecasts claiming that 15–37% of natural species will be 'committed to extinction' by 2050. An alternative is 'process-based' modelling, which aims to predict species distributions on the basis of resource allocation, demography or competition. They are theoretically more robust than niche-based models, but require much more ecological knowledge and data.

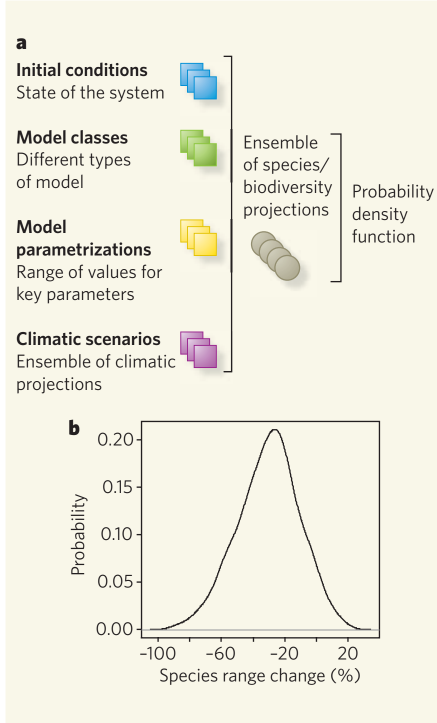


Figure 2 | The probabilistic approach to forecasting biodiversity. a, Each sub-activity produces an ensemble of projections based on subtly different initial conditions (such as factors influencing species distributions); on the class of model involved and its parametrization; and on the climate-change scenarios chosen. These ensembles are then combined to extract the possible range of outcomes and the likelihood of each occurring. Such estimates are called the 'probability density function' of the event being studied. **b,** An example of such a function, in which the projected range change of a given species is expressed as a probability of occurrence. In this case, there is an 80% probability that the given species will lose 20–60% of its current range. (Graphic based on M. B. Araújo & M. New *Trends Ecol. Evol.* 22, 42–47; 2007.)

What are the uncertainties behind forecasting?

All too many, starting with projections of climate change. It is no easy matter to accurately reflect complex interactions (such as those between the ocean and atmosphere), and account for different scenarios of greenhouse-gas emission. There is also our cruel lack of knowledge about the response of biota to rapid climate change. Few, if any, of the most popular models explicitly deal with migration, the dynamics at the trailing edge of shifting populations, species interactions, the interaction between the effects of climate and land use, and the direct effects of changes in atmospheric CO_2 and nitrogen deposition. At a basic level, ecologists are still debating the respective influence of interspecific competition and random events in shaping animal and plant communities. And different models tend to provide different predictions of species distribution or biodiversity under similar scenarios of environmental change, showing their limitations.

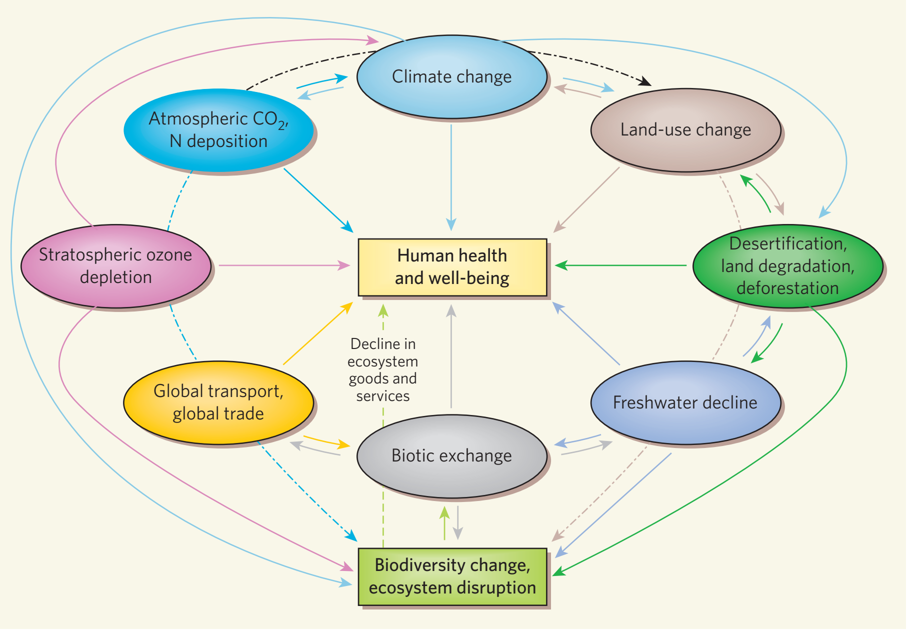


Figure 3 | The complex web of factors affecting human health and well-being, biodiversity and ecosystems. Changes in land use through land degradation, and climate change, are the most prominent factors. Perturbation of 'ecosystem goods and services' is just one part of this bigger picture.

Can forecasting be improved?

Large-scale, long-term experiments and observations are required to provide the data to make generalization possible, and for modelling studies. Mountains lend themselves to being natural laboratories, given that research can be carried out over steep gradients to investigate the differential response of species and the influence of local adaptations. Overall, what is needed is information that, when appropriately synthesized, can be applied to determine and fine-tune the parameters to be used in process-based models. The building of global databases is a big step forward in accumulating meta-information for this purpose. These databases include compilations of genetic sequences of species (for example, GenBank), the phylogenetic relationships of species (Tree of Life, Phylocom, TreeBASE), and measures of species traits such as mode of dispersal and competitive ability (TraitNet). There is also a new generation of hybrid models of species distributions, which aim for a compromise between realism and accuracy, and complexity and simplicity. These developments are opening up new ways to address the pressing ecological questions: combining hybrid models with statistical advances in 'ensemble' forecasting promises to provide probabilistic projections (Fig. 2).

What use are forecasts for conservation planning?

For all their imperfections, they are essential. For example, projections of species distributions guide the management of organisms under threat by helping to identify biological corridors for dispersal, sites for reintroduction and areas requiring protection. Lately, the conservation agenda has moved on to consider adaptation to climate change, and to test strategies such as habitat re-creation, creation of dispersal corridors and enhancing the resilience of ecosystems to changed conditions. An alternative is to identify desired future states, and then use models for 'back-casting' to identify strategies for achieving those states. Modellers need to explore how far species-distribution models can be taken to answer the crucial questions that arise from rapidly changing climate. Invasive species are a case in point. In principle, forecasts can predict the probability of an invasive species becoming established, and can incorporate early warning systems for controlling it.

How do human societies fit into this picture?

Much debate has centred on how climate change will affect human welfare through, for instance, rising sea levels and different patterns of crop production. But that well-being also depends on the diversity of organisms used for such 'ecosystem goods and services' as food, energy production and medicines. In certain parts of the world, the chain linking biodiversity, ecosystem processes, and ecosystem goods and services is likely to be broken as biodiversity is affected by altered climatic conditions and the many other factors affecting human health and well-being (Fig. 3). Here again, forecasting can be used to formulate policies that will ameliorate the consequences. For instance, forests are among the most valuable sources of ecosystem goods and services. 'Forest-gap' models can predict tree growth and biomass, the result then being used to guide forest conservation and production strategies.

What can we conclude from all this?

As outlined above, our ecological knowledge base and modelling capabilities are far from adequate: making swifter progress will depend on attracting the best scientific talent and the funds to work on these immensely intricate issues. That apart, forecasts of the consequences of climate change for biodiversity need to be couched in probabilistic terms, by stating the possible range of outcomes and estimating the likelihood of each occurring — as is now common practice in weather forecasting. That then presents the problem of recommending a particular course of action for particular circumstances. But if that step can be taken, we reach the stage at which action comes down to political will, at levels running from the global to the individual village. ■

Wilfried Thuiller is at the Laboratoire d'Ecologie Alpine, UMR-CNRS 5553, Université J. Fourier, BP 53, 38041 Grenoble Cedex 9, France.
e-mail: wilfried.thuiller@ujf-grenoble.fr

FURTHER READING

Díaz, S., Fargione, J., Chapin, F. S. III & Tilman, D. *PLoS Biol.* **4**(8), e277. doi:10.1371/journal. pbio.0040277 (2006).
Intergovernmental Panel on Climate Change *Climate Change 2007: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the IPCC* (Cambridge Univ. Press, 2007).
Millennium Ecosystem Assessment *Ecosystems and Human Well-Being: A Framework for Assessment* (Island, Washington DC, 2003).
Parmesan, C. *Annu. Rev. Ecol. Evol. Syst.* **37,** 637–669 (2006).
Sala, O. E. *et al.* in *Ecosystems and Human Well-Being: Scenarios* (eds Carpenter, S. R., Pingali, P. L., Bennett, E. M. & Zurek, M.) 375–408 (Island, Washington DC, 2005).
Schröter, D. *et al. Science* **310,** 1333–1337 (2005).
Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. *Proc. Natl Acad. Sci. USA* **102,** 8245–8250 (2005).

Vol 448 | 2 August 2007 | doi:10.1038/nature06008

ARTICLES

Genome-wide maps of chromatin state in pluripotent and lineage-committed cells

Tarjei S. Mikkelsen[1,2], Manching Ku[1,4], David B. Jaffe[1], Biju Issac[1,4], Erez Lieberman[1,2], Georgia Giannoukos[1], Pablo Alvarez[1], William Brockman[1], Tae-Kyung Kim[5], Richard P. Koche[1,2,4], William Lee[1], Eric Mendenhall[1,4], Aisling O'Donovan[4], Aviva Presser[1], Carsten Russ[1], Xiaohui Xie[1], Alexander Meissner[3], Marius Wernig[3], Rudolf Jaenisch[3], Chad Nusbaum[1], Eric S. Lander[1,3]* & Bradley E. Bernstein[1,4,6]*

We report the application of single-molecule-based sequencing technology for high-throughput profiling of histone modifications in mammalian cells. By obtaining over four billion bases of sequence from chromatin immunoprecipitated DNA, we generated genome-wide chromatin-state maps of mouse embryonic stem cells, neural progenitor cells and embryonic fibroblasts. We find that lysine 4 and lysine 27 trimethylation effectively discriminates genes that are expressed, poised for expression, or stably repressed, and therefore reflect cell state and lineage potential. Lysine 36 trimethylation marks primary coding and non-coding transcripts, facilitating gene annotation. Trimethylation of lysine 9 and lysine 20 is detected at satellite, telomeric and active long-terminal repeats, and can spread into proximal unique sequences. Lysine 4 and lysine 9 trimethylation marks imprinting control regions. Finally, we show that chromatin state can be read in an allele-specific manner by using single nucleotide polymorphisms. This study provides a framework for the application of comprehensive chromatin profiling towards characterization of diverse mammalian cell populations.

One of the fundamental mysteries of biology is the basis of cellular state. Although they have essentially identical genomes, the different cell types in a multicellular organism maintain markedly different behaviours that persist over extended periods. The most extreme case is lineage commitment during development, where cells progress from totipotency to pluripotency to terminal differentiation; each step involves establishment of a stable state encoding specific developmental commitments that can be faithfully transmitted to daughter cells. Considerable evidence suggests that cellular state may be closely related to 'chromatin state'—that is, modifications to histones and other proteins that package the genome[1–3]. Accordingly, it would be desirable to construct 'chromatin-state maps' for a wide variety of cell types, showing the genome-wide distribution of important chromatin modifications.

Chromatin state can be studied by chromatin immunoprecipitation (ChIP), in which an antibody is used to enrich DNA from genomic regions carrying a specific epitope. The major challenge to generating genome-wide chromatin-state maps lies in characterizing these enriched regions in a scalable manner. Enrichment at individual loci is commonly assayed by polymerase chain reaction (PCR), but this method does not scale efficiently. A more recent approach has been ChIP-chip, in which enriched DNA is hybridized to a microarray[4,5]. This technique has been successfully used to study large genomic regions. However, ChIP-chip suffers from inherent technical limitations: (1) it requires large amounts (several micrograms) of DNA and thus involves extensive amplification, which introduces bias; (2) it is subject to cross-hybridization, which hinders the study of repeated sequences and allelic variants; and (3) it is currently expensive to study entire mammalian genomes. Given these issues, only a handful of whole-genome ChIP-chip studies in mammals have been reported.

In principle, chromatin could be readily mapped across the genome by sequencing ChIP DNA and identifying regions that are over-represented among these sequences. Notably, sequence-based mapping could require relatively small quantities of DNA and provide nucleotide-level discrimination of similar sequences, thereby maximizing genome coverage. The major limitation has been that high-resolution mapping requires millions of sequences (Supplementary Note 1). This is cost-prohibitive with traditional technology, even with concatenation of multiple sequence tags[6]. However, recent advances in single-molecule-based sequencing (SMS) technology promise to increase throughput and decrease costs markedly[7]. In the approach developed by Illumina/Solexa, DNA molecules are arrayed across a surface, locally amplified, subjected to successive cycles of primer-mediated single-base extension (using fluorescently labelled reversible terminators) and imaged after each cycle to determine the inserted base. The 'read length' is short (25–50 bases), but tens of millions of DNA fragments may be read simultaneously.

Here, we report the development of a method for mapping ChIP enrichment by sequencing (ChIP-Seq) and describe its application to create chromatin-state maps for pluripotent and lineage-committed mouse cells. The resulting data define three broad categories of promoters based on their chromatin state in embryonic stem (ES) cells, including a larger than anticipated set of 'bivalent' promoters; reveal that lineage commitment is accompanied by characteristic chromatin changes at bivalent promoters that parallel changes in gene expression and transcriptional competence; and demonstrate the potential for using ChIP for genome-wide annotation of novel promoters and primary transcripts, active transposable elements, imprinting control regions and allele-specific transcription. This study provides a technological framework for comprehensive characterization of chromatin state across diverse mammalian cell populations.

[1]Broad Institute of Harvard and MIT, [2]Division of Health Sciences and Technology, MIT, and [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. [4]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [5]Department of Neurology, Children's Hospital, and [6]Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

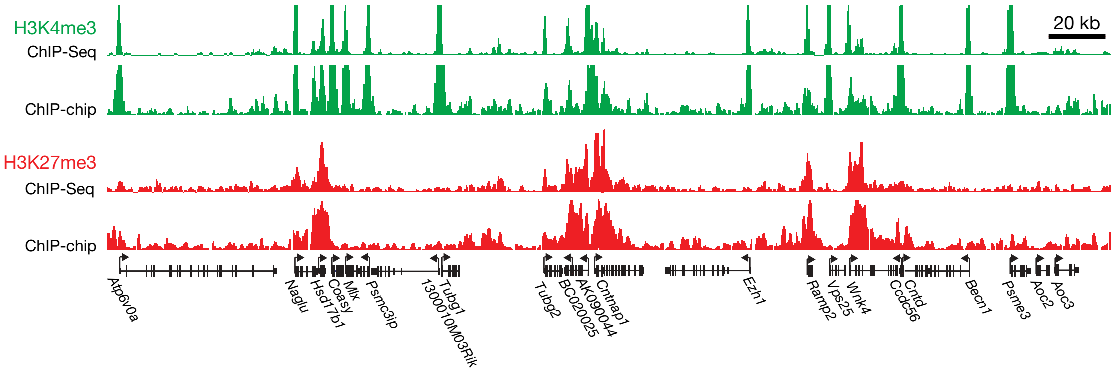


Figure 1 | Comparison of ChIP-Seq and ChIP-chip data. Direct comparison of H3K4me3 (green) and H3K27me3 (red) ChIP data across a 300-kb region in mouse ES cells from independent experiments assayed by SMS (absolute fragment counts) or tiling arrays (log *P*-values for enrichment relative to whole-cell extracts[15]).

Genome-wide chromatin-state maps

We created genome-wide chromatin-state maps for three mouse cell types: ES cells, neural progenitor cells (NPCs)[8] and embryonic fibroblasts (MEFs). For each cell type, we prepared and sequenced ChIP DNA samples for some or all of the following features: pan-H3, trimethylated histone H3 lysine 4 (H3K4me3), H3K9me3, H3K27me3, H3K36me3, H4K20me3 and RNA polymerase II (Supplementary Table 1).

In each case, we sequenced nanogram quantities of DNA fragments (~300 base pairs (bp)) on an Illumina/Solexa sequencer. We obtained an average of 10 million successful reads, consisting of the terminal 27–36 bases of each fragment. The reads were mapped to the genome and used to determine the number of ChIP fragments overlapping any given position (Fig. 1). Enriched intervals were defined as regions where this number exceeded a threshold defined by randomization (see Methods). The full data set consists of 18 chromatin-state maps, containing ~140 million uniquely aligned reads, representing over 4 billion bases of sequence.

We validated the chromatin-state maps by computational analysis and by comparison to previous methods. ChIP-Seq maps of specific histone modifications show marked enrichment at specific locations in the genome, whereas the pan-H3 and unenriched samples show relatively uniform distributions (Supplementary Figs 1 and 2). The maps show close agreement with our previously reported ChIP-chip data from ~2.5% of the mouse genome[9] (Fig. 1). Also, ChIP-PCR assays of 50 sites chosen to represent a range of ChIP-Seq fragment

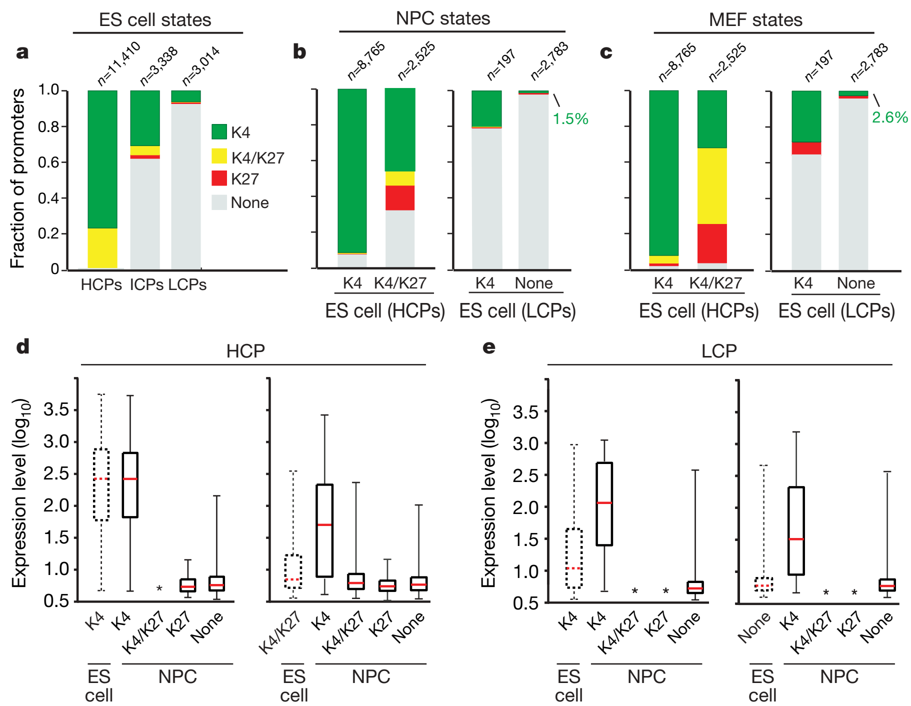


Figure 2 | Histone trimethylation state predicts expression of HCPs and LCPs. **a**, Mammalian promoters can be readily classified into sets with high (HCPs), intermediate (ICPs) or low (LCPs) CpG-content. In ES cells, virtually all HCPs are marked by H3K4me3, either alone (green) or in combination with H3K27me3 (yellow). In contrast, most LCPs have neither mark (grey). Few promoters are only enriched for H3K27me3 (red). **b**, Trimethylation states of HCPs and LCPs in NPCs (indicated by colours), conditional on their ES cell state (indicated below each bar). HCPs marked by H3K4me3 only in ES cells tend to retain this mark. HCPs marked by H3K4me3 and H3K27me3 tend to lose one or both marks, although some remain bivalent. Small, partially overlapping subsets of LCPs are marked by H3K4me3. **c**, Trimethylation states of HCPs and LCPs in MEFs. **d**, Changes in expression levels of HCP genes with H3K4me3 alone (left) or also with H3K27me3 (right) upon differentiation to NPCs. Resolution of bivalent promoters to H3K4me3 is associated with increased expression. **e**, Changes in expression levels of LCP genes with H3K4me3 (left) or no mark (right) upon differentiation to NPCs. Gain of H3K4me3 is associated with increased expression. For **d** and **e**, boxplots show median (red bar), 25th and 75th percentile expression levels in ES cells. Whiskers show 2.5th and 97.5th percentiles. Asterisks indicate classes with less than 15 genes.

counts showed 98% concordance and a strong, quantitative correlation (Supplementary Fig. 3 and Supplementary Table 2).

Promoter state in ES and lineage-committed cells

We began our analysis by studying H3K4me3 and H3K27me3 patterns at known promoters. H3K4me3 is catalysed by trithorax-group (trxG) proteins and associated with activation, whereas H3K27me3 is catalysed by polycomb-group (PcG) proteins and associated with silencing[10,11]. Recently, we and others observed that some promoters in ES cells carry both H3K4me3 and H3K27me3[9,12]. We termed this novel combination a 'bivalent' chromatin mark and proposed that it serves to poise key developmental genes for lineage-specific activation or repression.

We studied 17,762 promoters inferred from full-length transcripts (Supplementary Table 3). Mammalian RNA polymerase II promoters are known to occur in at least two major forms[13,14] (Supplementary Fig. 4). CpG-rich promoters are associated with both ubiquitously expressed 'housekeeping' genes, and genes with more complex expression patterns, particularly those expressed during embryonic development. CpG-poor promoters are generally associated with highly tissue-specific genes. Accordingly, we divided our analysis to focus on 'high' CpG promoters (HCP; $n = 11,410$) and 'low' CpG promoters (LCP; $n = 3,014$) separately. To ensure a clean separation, we defined a set of intermediate CpG content promoters (ICP; $n = 3,338$); this class shows properties consistent with being a mixture of the two major classes.

High CpG promoters in ES cells. Virtually all HCPs (99%) are associated with intervals of significant H3K4me3 enrichment in ES cells (Fig. 2a). The modified histones are typically confined to a punctate interval of 1–2 kilobases (kb) (Supplementary Fig. 5). As observed previously[15,16], there is a strong correlation between the intensity of H3K4me3 and the expression level of the associated genes (Spearman's $\rho = 0.67$). However, not all promoters associated with H3K4me3 are active.

The chromatin-state maps reveal that ~22% of HCPs ($n = 2,525$) are actually bivalent, exhibiting both H3K4me3 and H3K27me3 (Fig. 2a). A minority of these ($n = 564$) are 'wide' bivalent sites in which H3K27me3 extends over a region of at least 5 kb and resemble those described previously[9]. The majority ($n = 1,961$) are 'narrow' bivalent sites, with more punctate H3K27me3, that correspond to many additional PcG target promoters[17–19]. Bivalent promoters show low activity despite the presence of H3K4me3, suggesting that the repressive effect of PcG activity is generally dominant over the ubiquitous trxG activity (Supplementary Fig. 6 and Supplementary Table 4).

The different types of chromatin marks at HCPs are closely related to the nature of the associated genes (Supplementary Table 5). Monovalent promoters (H3K4me3) generally regulate genes with 'housekeeping' functions including replication and basic metabolism. By contrast, bivalent promoters are associated with genes with more complex expression patterns, including key developmental transcription factors, morphogens and cell surface molecules. In addition, several bivalent promoters appear to regulate transcripts for lineage-specific microRNAs.

High CpG promoters in NPCs and MEFs. Most HCPs marked with H3K4me3 alone in ES cells retain this mark both in NPCs and MEFs (92% in each; Figs 2b, c and 3a). This is consistent with the tendency for this sub-class of promoters to regulate ubiquitous housekeeping genes. A small proportion (~4%) of these promoters have H3K27me3 in MEFs, and are thus bivalent or marked by H3K27me3 alone. This correlates with lower expression levels and may reflect active recruitment of PcG proteins to new genes during differentiation[20]. An example is the transcription factor gene *Sox2*, where the promoter is marked by H3K4me3 alone in ES cells and NPCs, but H3K27me3 alone in MEFs. Notably, this locus is flanked by CpG islands with bivalent markings in ES cells (see below), suggesting that the locus may be poised for repression upon differentiation.

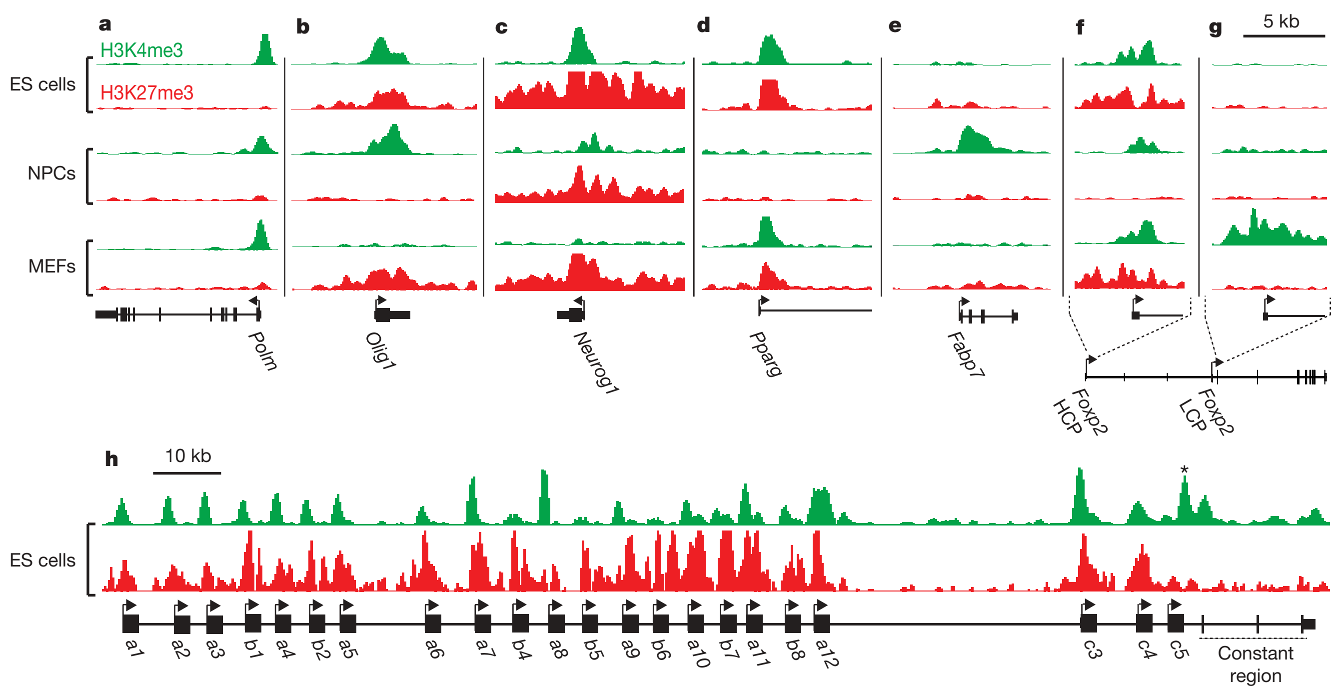


Figure 3 | Cell-type-specific chromatin marks at promoters. a, Multiple 'housekeeping genes', such as DNA polymerase μ (*Polm*), are associated with HCPs marked by H3K4me3 in all cell types. **b**, The neural transcription factor gene *Olig1* (HCP) is bivalent in ES cells, but resolves to H3K4me3 in NPCs and H3K27me3 in MEFs. **c**, The neurogenesis transcription factor gene *Neurog1* (HCP) remains bivalent upon differentiation to NPCs, but resolves to H3K27me3 in MEFs. **d**, The adipogenesis transcription factor gene *Pparg* (HCP) remains bivalent in MEFs, but loses both marks in NPCs. **e**, The neural progenitor marker gene *Fabp7* (LCP) is marked by H3K4me3 in NPCs only. **f**, The brain and lung expressed transcription factor gene *Foxp2* is associated with an HCP that is bivalent in ES cells, but resolves to H3K4me3 in NPCs and remains bivalent in MEFs. **g**, *Foxp2* also has an LCP marked by H3K4me3 in MEFs only. **h**, Multiple, distinct bivalent chromatin marks at the variable region promoters of *Pcdhg*. A promoter proximal to the constant region exons (asterisk) is marked by H3K4me3 only.

The majority of HCPs with bivalent marks in ES cells resolve to a monovalent status in the committed cells. In NPCs, 46% resolve to H3K4me3 only and these genes show increased expression (Figs 2b, d and 3b). Of the remaining promoters, 14% resolve to H3K27me3 alone and 32% lose both marks, with both outcomes being associated with low levels of expression. Notably, 8% remain bivalent and these genes also continue to be repressed (Figs 2b, d and 3c). A less resolved pattern is seen in MEFs, with 32% marked by H3K4me3 alone, 22% marked by H3K27me3 alone, 3% without both marks, and the remaining (43%) still bivalent (Fig. 2c). The relatively high number of bivalent promoters in MEFs may reflect a less differentiated state and/or heterogeneity in the population.

Distinct regulation of low CpG promoters. The LCPs show a very different pattern than the HCPs. Only a small minority (6.5%, $n = 207$) of LCPs have significant H3K4me3 in ES cells and virtually none have H3K27me3 (Fig. 2a). Most of these promoters have lost H3K4me3 in NPCs and MEFs, whereas a small number of other LCPs (1.5% and 2.6%, respectively) have gained the mark (Figs 2b, c and 3e). In all three cell types, the expression levels of the associated genes strongly correlate with the presence or absence of H3K4me3 (Fig. 2e and Supplementary Fig. 6).

The genes with LCPs marked by H3K4me3 are closely related to tissue-specific functions. In NPCs, they include genes encoding several known markers of neural progenitors *in vivo* (such as *Fabp7*, *Cp*, *Gpr56*). In MEFs, they include genes encoding extracellular matrix components and growth factors (such as *Col3a1*, *Col6a1*, *Postn*, *Aspn*, *Hgf*, *Figf*), consistent with the mesenchymal origin of these cells (see below).

We conclude that HCPs and LCPs are subject to distinct modes of regulation. In ES cells, all HCPs seem to be targets of trxG activity, and may therefore drive transcription unless actively repressed by PcG proteins. In committed cell types, a subset of HCPs appears to lose the capacity to recruit trxG activity (possibly due to other epigenetic modifications, such as DNA methylation[21]). In contrast, CpG-poor promoters seem to be inactive by default, independent of repression by PcG proteins, and may instead be selectively activated by cell-type- or tissue-specific factors.

Alternative promoter use. We note that genes with alternative promoters may have multiple, distinct chromatin states. An 'active' state at any one of these promoters may be sufficient to drive expression. A common situation involves genes with one major HCP and one or more alternative LCPs. An example is the transcription factor *Foxp2*, which is expressed at moderate levels in both NPCs and MEFs (Fig. 3f, g). The *Foxp2* HCP is marked by H3K4me3 in NPCs, but is bivalent in MEFs. However, an alternative LCP is marked by H3K4me3 exclusively in MEFs. The protocadherin-γ (*Pcdhg*) locus is a more extreme case: the amino-terminal variable regions of this gene are transcribed from at least 20 different HCPs in neurons[22], all of which carry bivalent chromatin marks in ES cells. *Pcdhg* expression is nevertheless detected by microarrays, possibly owing to a single promoter in front of the carboxy-terminal constant region marked by H3K4me3 alone (Fig. 3h).

Although only ~10% of the genes analysed here have more than one known promoter, recent 'cap-trapping' studies suggest that alternative promoter use may be substantially more common[23]. The ability of ChIP-Seq to assess chromatin state at known promoters, as well as to identify novel promoters (see below), should prove valuable in analysis of transcriptional networks.

Promoter state reflects lineage commitment and potential

Given their association with epigenetic memory, we next examined whether the patterns of H3K4me3 and H3K27me3 can reflect developmental potential. Both of the committed cell types studied here have been shown to be multipotent *ex vivo*. NPCs can be differentiated to glial and neuronal lineages[8], whereas primary MEFs have been differentiated into adipocytes[24], chondrocytes[25] and osteoblast-like cells[26].

Lineage-specific resolution and retention of bivalent marks. We first examined a set of genes involved in *in vivo* differentiation pathways known to be, at least partially, recapitulated by MEFs, NPCs, or neither. These genes all have bivalent promoters in ES cells. We found that their resolution in lineage-committed cells is closely related to their demonstrated developmental potential (Supplementary Table 6): (1) genes restricted to regulation or specialized functions in unrelated lineages, such as haematopoietic (*Cdx4*, *PU.1* (also called *Sfpi1*)), epithelial (*Cnfn*, *Krt2-4*), endoderm (*Gata6*, *Pdx1*) or germ line (*Tenr* (*Adad1*), *Ctcfl*), generally resolved to monovalent H3K27me3 or carry neither mark in both NPCs and MEFs. (2) Genes related to adipogenesis and chondro/osteogenesis often remain bivalent in MEFs, but not in NPCs. Examples include *Pparg*, which is a key regulator of adipogenesis, and *Sp7*, which promotes chondro/osteogenic pathways. Early mesenchymal markers, such as *Runx1* and *Sox9*, resolved to H3K4me3 alone in MEFs. (3) Genes related to gliogenesis and neurogenesis often resolved to H3K4me3 alone or remain bivalent in NPCs, whereas they resolved to H3K27me3 alone in the MEFs. Gliogenesis and neurogenesis are thought to be mutually opposing pathways[27], and we find that genes promoting gliogenesis are more likely to resolve to H3K4me3 in NPCs. Examples include *Bmp2* and the microRNA *mir-9-3*, which promotes glial but inhibits neuronal differentiation[28]. Several genes known to promote neuronal differentiation, such as *Neurog1* and *Neurog2*, remain bivalent whereas others, such as *Bmp6*, appear to resolve to H3K27me3 alone. In our hands, the NPCs differentiate to astrocytes with significantly higher efficiency than to neurons (M.W., unpublished data). The observed chromatin patterns may reflect this gliogenic bias.

Correlation with expression in adult tissues. We next analysed gene expression in adult tissues with major contributions from neuroectodermal or mesenchymal lineages. We reasoned that if H3K4me3 is generally not restored once lost, then differential loss of H3K4me3 at promoters early in these lineages (as represented by NPCs and MEFs, respectively) might be reflected in differential gene expression patterns in related adult tissues.

Notably, we observed a clear bias in relative expression levels between relevant adult tissues for genes that retain H3K4me3 in NPCs only versus genes that retain H3K4me3 in MEFs only. The former are strongly biased towards higher expression in various brain sections, whereas the latter are biased towards higher expression in bone, adipose and other mesenchyme-rich tissues (Fig. 4).

These analyses are limited by alternative promoter usage, the cell models used, and the heterogeneity of the adult tissues. Nonetheless, the data show clear trends that support an important role for retention and resolution of bivalent chromatin in the regulation of hierarchical lineage commitment.

Genome-wide annotation of promoters and primary transcripts

We next considered genome-wide maps of H3K36me3. This mark has been linked to transcriptional elongation and may serve to prevent aberrant initiation within gene bodies[29–33]. Our chromatin maps reveal a global pattern of H3K36me3 in mammals similar to that previously observed in yeast[29].

In all three cell types, H3K36me3 is strongly enriched across the transcribed regions of active genes (Fig. 5a), beginning immediately after the promoter H3K4me3 signal. The level of H3K36me3 is strongly correlated with the level of gene expression (Spearman's $\rho = 0.77$), although the dynamic range is compressed (1–2 orders of magnitude for H3K36me3 versus 3–4 for expression levels; Supplementary Fig. 7). Genes with bivalent promoters rarely show H3K36me3, consistent with their low expression. Notably, there is essentially no overlap between intervals significantly enriched for H3K36me3 and for H3K27me3, consistent with a role for PcG complexes in the exclusion of polymerases[11].

The vast majority of intervals significantly enriched for H3K36me3 is associated with known genes (~92% in ES cells), but there are

at least ~500 additional regions across the genome (median size ~2 kb), with most being adjacent to sites of H3K4me3. Inspection revealed a number of interesting cases, falling into three categories.

The first category corresponds to H3K36me3 that extends significantly upstream from the annotated start of a known gene, often until an H3K4me3 site. These seem to reflect the presence of unannotated alternative promoters. A notable example is the *Foxp1* locus. In ES cells, one annotated *Foxp1* promoter is marked by H3K4me3 and another CpG-rich region located ~500 kb upstream carries a bivalent mark. In MEFs, this CpG island is marked by H3K4me3 only, and H3K36me3 extends from this site to the 3′ end of *Foxp1* (Fig. 5a). Although no transcript extending across this entire region has been reported in mouse, the orthologous position in human has been shown to act as a promoter for the orthologous gene. The ChIP-Seq data contain many other examples where the combination of H3K36me3 and H3K4me3 seems to reveal novel promoters.

The second category corresponds to H3K36me3 that extends significantly downstream of a known gene. An example is the *Sox2* locus, which encodes a pluripotency-associated transcription factor that also functions during neural development. In ES cells, *Sox2* has an unusually large region of H3K4me3 (>20 kb) accompanied by H3K36me3 extending far beyond the annotated 3′ end (>15 kb); non-coding transcription throughout the locus has been noted previously[34] and may serve a regulatory role (Fig. 5b).

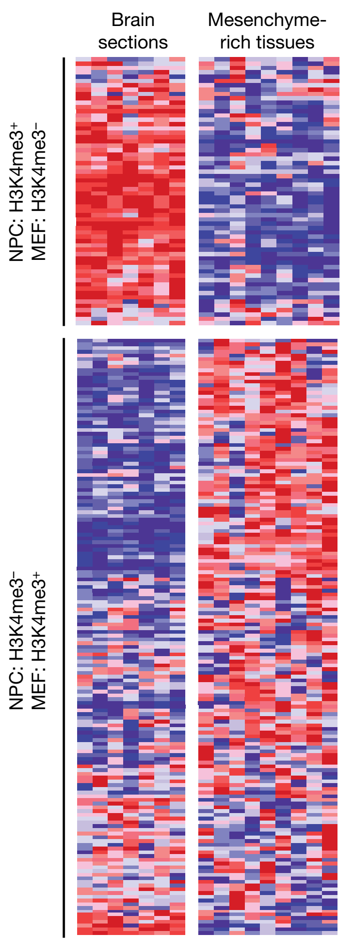


Figure 4 | Correlation between chromatin-state changes and lineage expression. Relative expression levels across adult mouse brain (frontal and cerebral cortex, substantia nigra, cerebellum, amygdale, hypothalamus, hippocampus) and relatively mesenchyme-rich tissues (bone, white fat, brown fat, trachea, digits, lung, bladder, uterus, umbilical cord) are shown for genes with bivalent chromatin marks in ES cells that retain H3K4me3 in NPCs but lose this mark in MEFs ($n = 62$) or vice versa ($n = 160$). Red, white and blue indicate higher, equal and lower relative expression, respectively.

The third category seems to reflect transcription of non-coding RNA genes. For example, two regions with H3K36me3 and adjacent H3K4me3 correspond to recently discovered nuclear transcripts with possible functions in messenger RNA processing[35] (Fig. 5c). In addition, a number of these presumptive transcriptional units overlap microRNAs (Fig. 5d). A striking example is a >200-kb interval within the *Dlk1–Dio3* imprinted locus (Fig. 6a). This region harbours over 40 non-coding RNAs, including clusters of microRNAs and small nucleolar RNAs[36]. The ChIP-Seq data suggest that the entire region is transcribed as a single unit that initiates at a H3K4me3-marked HCP.

These findings suggest that genome-wide maps of H3K4me3 and H3K36me3 may provide a general tool for defining novel transcription units. The capacity to define the origins and extents of primary transcripts will be of particular value for characterizing the regulation of microRNAs and other non-coding RNAs that are rapidly processed from long precursors[37]. Finally, the relatively narrow dynamic range of H3K36me3 may offer advantages over RNA-based approaches in assessing gene expression and defining cellular states.

H3K9 and H4K20 trimethylation mark specific repetitive elements

We next studied H3K9me3 and H4K20me3, both of which have been associated with silencing of centromeres, transposons and tandem repeats[38–40]. We sought first to assess the relative enrichments of H3K9me3 and H4K20me3 across different types of repetitive elements by aligning ChIP-Seq reads directly to consensus sequences for various repeat families (~40 million reads could be aligned this way).

H3K9me3 and H4K20me3 show nearly identical patterns of enrichment in ES cells. The strongest enrichments are observed for telomeric, satellite and long terminal repeats (LTRs). The LTR signal

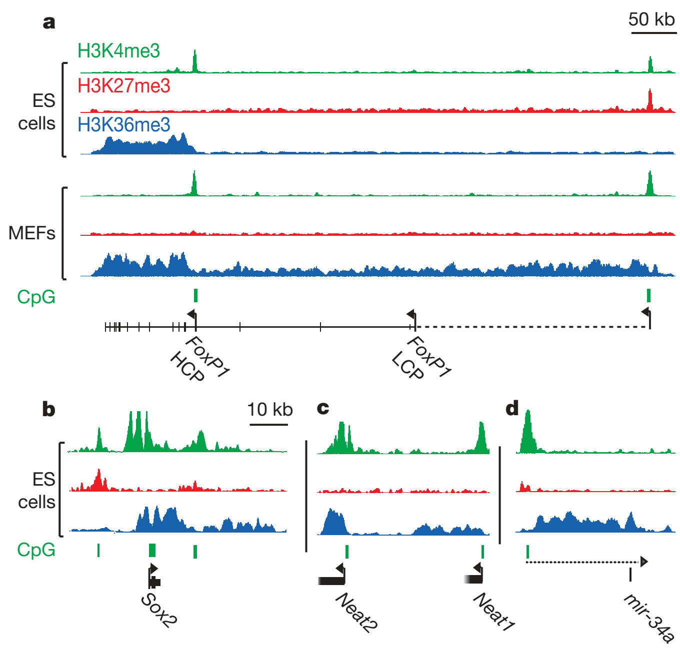


Figure 5 | H3K4me3 and H3K36me3 annotate genes and non-coding RNA transcripts. **a**, *Foxp1* has two annotated promoters (based on RefSeq and UCSC 'known genes'), only one of which shows H3K4me3 in ES cells. The corresponding transcriptional unit is marked by H3K36me3. In MEFs, H3K36me3 extends an additional 500 kb upstream to an H3K4me3 site that seems to reflect an alternative promoter (this site is bivalent in ES cells). **b**, H3K36me3 enrichment extends significantly downstream of *Sox2*. Although highly active in ES cells, *Sox2* is flanked by two bivalent CpG islands that may poise it for repression. **c**, **d**, H3K4me3 and H3K36me3 indicate two highly expressed non-coding RNAs (**c**), and the putative primary transcript (dashed line) for a single annotated microRNA (**d**).

primarily reflects enrichment of intracisternal A-particle (IAP) and early transposon (ETn) elements (Supplementary Fig. 8).

IAP and ETn elements are active in murine ES cells and produce double-stranded RNAs[41,42]. RNA has also been implicated in maintaining satellite and telomeric heterochromatin[38]. Hence, these enrichment data are consistent with a global role for RNA in targeting repressive chromatin marks in mammalian ES cells, analogous to that observed in lower eukaryotes[38,39].

We next examined the distributions of H3K9me3 and H4K20me3 across unique sequences in the mouse genome. We identified ~1,800 H3K9me3 sites (median size ~300 bp) in ES cells, with the vast majority also showing H4K20me3. Fully 78% of the sites lie within 2 kb of a satellite repeat or LTR (primarily IAP and ETn elements). This suggests that repressive marks are capable of spreading from repeat insertions and could potentially regulate proximal unique sequence.

Recent studies have described a handful of active genes with H3K9me3 and H4K20me3, raising the possibility that these 'repressive' marks also function in transcriptional activation[31,32]. One-third of the ~1,800 H3K9me3-enriched sites reside within an annotated gene, which is roughly the proportion expected by chance. However, H3K9me3 sites that are larger and/or more distant from LTRs are more likely to occur within genes (Supplementary Fig. 9). The largest genic site in ES cells (~6 kb) coincides with the *Polrmt* gene (Fig. 6d). This case is notable because the downstream gene (*Hcn2*) is convergent and contains a CpG island at its 3′ end. Transcription from 3′ promoters has been proposed as a potential mechanism of transcriptional interference by producing antisense transcripts[23]. This example may therefore reflect a link between transcriptional interference and H3K9me3, as has been suggested for a few other mammalian loci[43,44]. Our results thus confirm the presence of H3K9me3 within a subset of genes, although the functional implications remain to be elucidated.

H3K4 and H3K9 trimethylation at imprinted loci

We next studied chromatin marks associated with imprinting. This epigenetic process typically involves allele-specific DNA methylation of CpG-rich imprinting control regions[45]. Several reports have also described allele-specific chromatin modification at a handful of imprinting control regions, with H3K9me3 and H4K20me3 on the DNA methylated allele and H3K4me3 on the opposite allele[46,47].

We searched for regions showing overlapping H3K9me3 and H3K4me3 in ES cells. Notably, 13 of the top 20 sites, as ranked by enrichment of the two marks, are located within known imprinted regions, coincident with imprinting control regions or imprinted gene promoters. An example is the *Peg13* promoter (Fig. 6c). Conversely, of the ~20 known and putative autosomal imprinted loci that contain imprinting control regions, 17 have at least one with the overlapping chromatin marks (Supplementary Table 7). We conclude that overlapping H3K9me3 and H3K4me3 is a common signature of imprinting control regions in ES cells.

Allele-specific histone methylation

To explore the feasibility of inferring allele-specific chromatin states, we constructed chromatin-state maps in male ES cells derived from a more distant cross (129SvJae (maternal) x *Mus musculus castaneus* (paternal)), and used a catalogue of ~3.5 million single nucleotide polymorphisms (SNPs) to assign ChIP-Seq reads to one of the two parental alleles.

As a positive control, we first compared results for chromosome X and the autosomes for reads derived by H3K4me3 ChIP. Virtually all

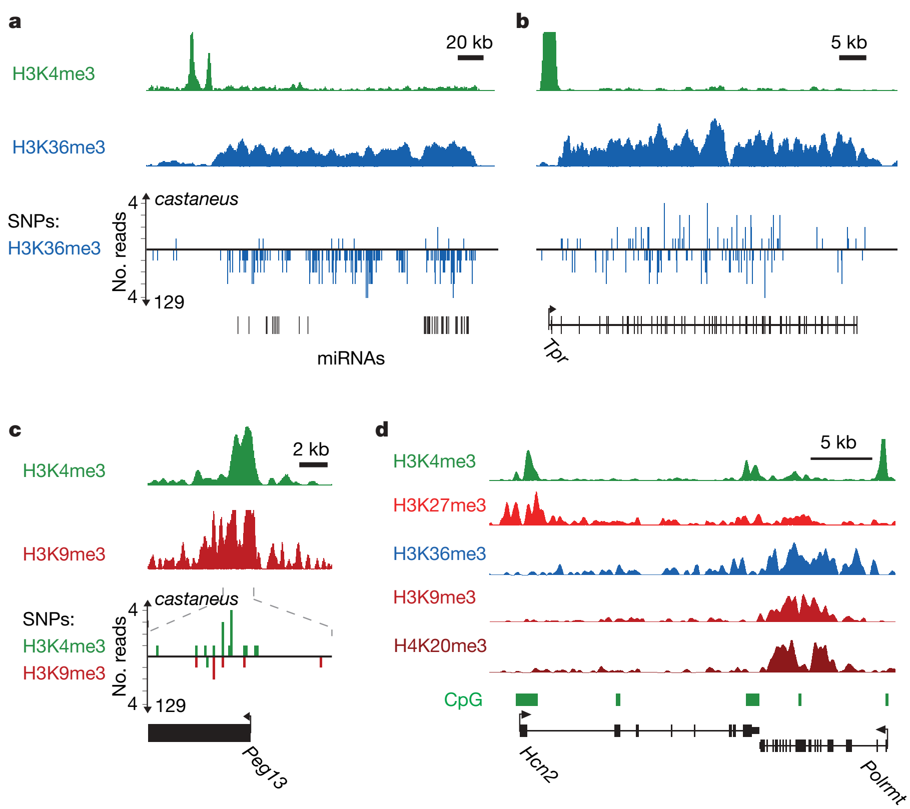


Figure 6 | Allele-specific histone methylation and genic H3K9me3/H4K20me3. **a**, H3K4me3 and H3K36me3 indicate a primary microRNA transcript in the *Dlk1–Dio3* locus. The allele specificity of this transcript is read out using ChIP-Seq data for hybrid ES cells and a SNP catalogue. The H3K36me3 reads overwhelmingly correspond to maternal 129SvJae alleles, consistent with the known maternal expression of these microRNAs[36]. **b**, In contrast, a non-imprinted transcript shows roughly equal proportions of reads assigned to 129SvJae and *M. m. castaneus* alleles. **c**, *Peg13* is marked by H3K4me3 and H3K9me3 in ES cells; 19 out of 21 H3K4me3 reads correspond to the paternal *M. m. castaneus* allele, whereas 6 out of 6 H3K9me3 reads correspond to the maternal 129SvJae allele, consistent with paternal expression of this gene. **d**, H3K9me3 and H4K20me3 enrichment evident at the *Polrmt* gene may reflect transcriptional interference owing to antisense transcription from the 3′ UTR CpG island of *Hcn2* (see text).

(97%) of ~3,700 informative reads on chromosome X, and roughly half (57%) of the 178,000 informative reads on the autosomes, were assigned to the 129SvJae strain. These proportions correspond roughly to the expected 100% and 50%.

We then examined the allelic distribution at overlapping H3K4me3 and H3K9me3 sites coincident with putative imprinting control regions (see above). Six of the imprinting control regions had enough reads (≥10) containing SNPs to assess allelic bias. In every case, the SNPs showed significant bias in the expected direction ($P < 0.02$; Fig. 6c and Supplementary Table 7).

We applied the same approach to search for allelic imbalance in intervals with significant H3K36me3 enrichment, which would predict differential transcription of the two alleles. A striking interval corresponds to a microRNA cluster within the *Dlk1–Dio3* locus known to be imprinted in the embryo proper[36] (Fig. 6a, b). Of the additional imprinted genes with H3K36me3 enrichment, four (*Snrpn*, *Grb10*, *Impact*, *Peg3*) had enough reads containing SNPs to assess allelic bias. In every case, the data showed significant bias in the expected direction ($P < 0.02$). The data also revealed novel instances of allele-specific transcription. For example, a transcript of unknown function (*BC054101*), first identified in trophoblast stem cells[48], showed highly significant maternal bias for H3K36me3, as well as H3K4me3 ($P < 10^{-15}$; Supplementary Fig. 10).

The results suggest that, with sufficiently deep coverage and dense SNP maps, ChIP-Seq will provide a powerful means for identifying allele-specific chromatin modifications. With data from reciprocal crosses, it should be possible to discriminate novel cases of imprinting from strain-specific differences.

Discussion

Genome-wide chromatin-state maps provide a rich source of information about cellular state, yielding insights beyond what is typically obtained by RNA expression profiling. Analysis of H3K4me3 and H3K36me3 allows recognition of promoters together with their complete transcription units. This should help to define alternative promoters and their usage in specific cell types; identify the primary structure of genes encoding non-coding RNAs; detect gene expression (given the narrower dynamic range); and detect allele-specific transcription. In addition, analysis of H3K9me3 and H4K20me3 should facilitate the study of heterochromatin, spreading and imprinting mechanisms.

Most interestingly, analysis of H3K4me3 and H3K27me3 provides a rich description of cellular state. Our results suggest that promoters may be classified as active, repressed or poised for alternative developmental fates. Conceivably, chromatin state at key regulatory genes may suffice to describe developmental commitment and potential.

Given the technical features of ChIP-Seq (high throughput, low cost and input requirement), it is now appropriate to contemplate projects to generate catalogues of chromatin-state maps representing a wide range of human and mouse cell types. These should include varied developmental stages and lineages, from totipotent to terminally differentiated, with the aim of precisely defining cellular states at the epigenetic level and observing how they change over the course of normal development. Chromatin-state maps should also be systematically catalogued from situations of abnormal development. Cancer cells are the most obvious targets, as they are frequently associated with epigenetic defects and many appear to have acquired characteristics of earlier developmental stages. A comprehensive public database of chromatin-state maps would be a valuable resource for the scientific community.

METHODS SUMMARY

Murine V6.5 ES cells (129SvJae × C57BL/6; male), hybrid ES cells (129SvJae × *M. m. castaneus* F_1; male) and NPCs were cultured as described[8,9]. Primary MEFs (129SvJae × C57BL/6; male) were obtained at embryonic day (E)13.5.

ChIP experiments were carried out as described[15]. Sequencing libraries were generated from 1–10 ng of ChIP DNA by adaptor ligation, gel purification and 18 cycles of PCR. Sequencing was carried out using the Illumina/Solexa Genome Analyzer system according to the manufacturer's specifications.

Reads were aligned to the reference genome, and the fragment count at any given position (25-bp resolution) was estimated as the number of uniquely aligned reads oriented towards it and within 300 bp. Enriched intervals were identified by comparison of the mean fragment count in 1-kb windows against a sample-specific expected distribution estimated by randomization (H3K4me3, H3K27me3), or using a supervised Hidden Markov Model (H3K36me3, H3K9me3, H4K20me3).

Promoters were inferred from full-length mouse RefSeqs. HCPs contain a 500-bp interval within −0.5 kb to +2 kb with a (G+C)-fraction ≥0.55 and a CpG observed to expected ratio (O/E) ≥0.6. LCPs contain no 500-bp interval with CpG O/E ≥0.4. Chromatin states of promoters were determined by overlap with H3K4me3- and H3K27me3-enriched intervals. Correlations with expression levels were calculated from the mean fragment count over each promoter or transcript.

ES cell, NPC and MEF expression data were generated using GeneChip arrays (Affymetrix) and GenePattern (http://www.broad.mit.edu/cancer/software/). Expression data for adult tissues were downloaded from Novartis (http://symatlas.gnf.org).

Repeat class enrichments were determined by aligning reads to consensus sequences (http://www.girinst.org). Mouse SNP maps were obtained from Perlegen (http://mouse.perlegen.com). Allele-specific bias was evaluated by a binomial test of the null hypothesis that ChIP fragments were drawn uniformly from both alleles.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 May; accepted 13 June 2007.
Published online 1 July 2007.

1. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
2. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
3. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
4. Buck, M. J. & Lieb, J. D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
5. Mockler, T. C. *et al.* Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1–15 (2005).
6. Roh, T. Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).
7. Service, R. F. Gene sequencing. The race for the $1000 genome. *Science* **311**, 1544–1546 (2006).
8. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283 (2005).
9. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
10. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
11. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
12. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532–538 (2006).
13. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103**, 1412–1417 (2006).
14. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
15. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
16. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
17. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
18. Lee, T. I. *et al.* Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
19. Squazzo, S. L. *et al.* Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res.* **16**, 890–900 (2006).
20. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The Polycomb Group protein Suz12 is required for Embryonic Stem Cell differentiation. *Mol. Cell. Biol.* **27**, 3769–3779 (2007).
21. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).

22. Wang, X., Su, H. & Bradley, A. Molecular mechanisms governing *Pcdh-γ* gene expression: evidence for a multiple promoter and *cis*-alternative splicing model. *Genes Dev.* **16,** 1890–1905 (2002).
23. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309,** 1559–1563 (2005).
24. Alexander, D. L., Ganem, L. G., Fernandez-Salguero, P., Gonzalez, F. & Jefcoate, C. R. Aryl-hydrocarbon receptor is an inhibitory regulator of lipid synthesis and of commitment to adipogenesis. *J. Cell Sci.* **111,** 3311–3322 (1998).
25. Lengner, C. J. *et al.* Primary mouse embryonic fibroblasts: a model of mesenchymal cartilage formation. *J. Cell. Physiol.* **200,** 327–333 (2004).
26. Garreta, E., Genove, E., Borros, S. & Semino, C. E. Osteogenic differentiation of mouse embryonic stem cells and mouse embryonic fibroblasts in a three-dimensional self-assembling peptide scaffold. *Tissue Eng.* **12,** 2215–2227 (2006).
27. Doetsch, F. The glial identity of neural stem cells. *Nature Neurosci.* **6,** 1127–1134 (2003).
28. Krichevsky, A. M., Sonntag, K. C., Isacson, O. & Kosik, K. S. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem Cells* **24,** 857–864 (2006).
29. Rao, B., Shibata, Y., Strahl, B. D. & Lieb, J. D. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* **25,** 9447–9459 (2005).
30. Bannister, A. J. *et al.* Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280,** 17732–17736 (2005).
31. Kim, A., Kiefer, C. M. & Dean, A. Distinctive signatures of histone methylation in transcribed coding and noncoding human β-globin sequences. *Mol. Cell. Biol.* **27,** 1271–1279 (2007).
32. Vakoc, C. R., Sachdeva, M. M., Wang, H. & Blobel, G. A. Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* **26,** 9185–9195 (2006).
33. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128,** 707–719 (2007).
34. Fantes, J. *et al.* Mutations in *SOX2* cause anophthalmia. *Nature Genet.* **33,** 461–463 (2003).
35. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8,** 39 (2007).
36. Seitz, H. *et al.* A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.* **14,** 1741–1748 (2004).
37. Cullen, B. R. Transcription and processing of human microRNA precursors. *Mol. Cell* **16,** 861–865 (2004).
38. Zaratiegui, M., Irvine, D. V. & Martienssen, R. A. Noncoding RNAs and gene silencing. *Cell* **128,** 763–776 (2007).
39. Verdel, A. & Moazed, D. RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett.* **579,** 5872–5878 (2005).
40. Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24,** 800–812 (2005).
41. Baust, C. *et al.* Structure and expression of mobile ETnII retroelements and their coding-competent MusD relatives in the mouse. *J. Virol.* **77,** 11448–11458 (2003).
42. Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269,** 276–285 (2004).
43. Cho, D. H. *et al.* Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20,** 483–489 (2005).
44. Feng, Y. Q. *et al.* The human β-globin locus control region can silence as well as activate gene expression. *Mol. Cell. Biol.* **25,** 3864–3874 (2005).
45. Edwards, C. A. & Ferguson-Smith, A. C. Mechanisms regulating imprinted genes in clusters. *Curr. Opin. Cell Biol.* **19,** 281–289 (2007).
46. Delaval, K. *et al.* Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* **26,** 720–729 (2007).
47. Feil, R. & Berger, F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.* **23,** 192–199 (2007).
48. Strausberg, R. L. *et al.* Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA* **99,** 16899–16903 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Fisher, M. Kellis, B. Birren and M. Zody for technical assistance and constructive discussions. We acknowledge L. Zagachin in the MGH Nucleic Acid Quantitation core for assistance with real-time PCR. E.M. was supported by an institutional training grant from NIH. M.W. was supported by fellowships from the Human Frontiers Science Organization Program and the Ellison Foundation. This research was supported by funds from the National Human Genome Research Institute, the National Cancer Institute, the Burroughs Wellcome Fund, Massachusetts General Hospital, and the Broad Institute of MIT and Harvard.

Author Information All analysed data sets can be obtained from http://www.broad.mit.edu/seq_platform/chip/. Microarray data have been submitted to the GEO repository under accession number GSE8024. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.S.L. (lander@broad.mit.edu) or B.E.B. (bbernstein@partners.org).

METHODS

Cell culture. V6.5 murine ES cells (genotype 129SvJae × C57BL/6; male; passages 10–15) and hybrid murine ES cells (genotype 129SvJae × *M. m. castaneus* F_1; male; passages 4–6) were cultivated in 5% CO_2 at 37 °C on irradiated MEFs in DMEM containing 15% FCS, leukaemia-inhibiting factor, penicillin/streptomycin, L-glutamine, nonessential amino acids and 2-mercaptoethanol. Cells were subjected to at least two to three passages on 0.2% gelatin under feeder-free conditions to exclude feeder contamination. V6.5 ES cells were differentiated into neural progenitor cells (NPCs) through embryoid body formation for 4 days and selection in ITSFn media for 5–7 days, and maintained in FGF2 and EGF2 (R&D Systems) as described[8]. The cells uniformly express nestin and *Sox2* and can differentiate into neurons, astrocytes and oligodendrocytes. Mouse embryonic fibroblasts (genotype 129SvJae × C57BL/6; male; E13.5; passages 4–6), were grown in DMEM with 10% fetal bovine serum and penicillin/streptomycin at 37°, 5% CO_2.

Chromatin immunoprecipitation. ChIP experiments were carried out as described previously[15] and at http://www.upstate.com. Briefly, chromatin from fixed cells was fragmented to a size range of 200–700 bases with a Branson 250 Sonifier or a Diagenode Bioruptor. Solubilized chromatin was immunoprecipitated with antibody against H3K4me3 (Abcam 8580), H3K9me3 (Abcam 8898), H3K27me3 (Upstate 07-449), H3K36me3 (Abcam 9050), H4K20me3 (Upstate 07-463), pan-H3 (Abcam 1791) or RNA polymerase II (Covance MMS-126R). Antibody–chromatin complexes were pulled-down using protein A-sepharose (or anti-IgM-conjugated agarose for RNA polymerase II), washed and then eluted. After cross-link reversal and proteinase K treatment, immunoprecipitated DNA was extracted with phenol-chloroform, ethanol precipitated, and treated with RNase. ChIP DNA was quantified using PicoGreen.

Library preparation and Solexa sequencing. One to ten nanograms of ChIP DNA (or unenriched whole-cell extract) were prepared for Solexa sequencing as follows: DNA fragments were repaired to blunt ends by T4 DNA polymerase and phosphorylated with T4 polynucleotide kinase using the END-IT kit (Epicentre). Then, a single 'A' base was added to 3′ ends with Klenow (3′→5′ exo^{-}, 0.3 U μl^{-1}). Double-stranded Solexa adaptors (75 bp with a 'T' overhang) were ligated to the fragments with DNA ligase (0.05 U μl^{-1}). Ligation products between 275 and 700 bp were gel purified on 2% agarose to remove unligated adaptors, and subjected to 18 PCR cycles. Completed libraries were quantified with PicoGreen.

DNA sequencing was carried out using the Illumina/Solexa Genome Analyzer sequencing system. Cluster amplification, linearization, blocking and sequencing primer reagents were provided in the Solexa Cluster Amplification kits and were used according to the manufacturer's specifications as described here. To obtain single strand templates, the sample preparation was first denatured in NaOH (0.1 N final concentration) and diluted in Solexa hybridization buffer (4 °C) to a final concentration of either 2 or 4 pM. Sample loading was carried out as follows. A template sample was loaded into each lane of a Solexa flowcell mounted on a Solexa cluster station on which all subsequent steps were performed. The temperature was increased to 95 °C for 1 min and slowly decreased to 40 °C to allow for annealing onto complementary adaptor oligonucleotides on the flowcell surface. Cluster formation was then carried out as follows. The template strands were extended with Taq polymerase (0.25 U μl^{-1}) to generate a fixed copy of the template on the flowcell. The samples were then denatured with formamide (Sigma-Aldrich, F-5786, >99.5% (GC)) and washed (Solexa Wash buffer) to remove the original captured template, leaving behind a single-stranded template ready for amplification. Clusters were then amplified under isothermal conditions (60 °C) for 30 cycles using Solexa amplification mix containing *Bst*I DNA polymerase (0.08 U μl^{-1}). After each amplification cycle, the templates were denatured with formamide (as above). Fresh amplification mix was added after each denaturation step. After amplification, the clusters were linearized with Solexa linearization mix, and any unextended flowcell surface capture oligonucleotides were blocked with ddNTPs (2.4 μM mix in the presence of 0.25 U μl^{-1} terminal transferase). The linearized clusters were then denatured (0.1 N NaOH) to remove and wash away the linearized strands. The single-stranded templates in the cluster were then annealed with the Solexa sequencing primer (10 μM). The flowcells were removed from the cluster station and then transferred onto the 1G Genetic Analyser which performed the sequencing according to its own standard protocols. We followed the protocol without any modifications.

Read alignment and generation of density maps and modified intervals. Sequence reads from each ChIP library are compiled, post-processed and aligned to the reference genome sequence using a general purpose computational pipeline. We first pre-compute a table that associates each possible 12-mer with all of its occurrences in the reference genome. Then, for each SMS read, we scan both it and its reverse complement, and for each of its constituent 12-mers, we find each potential start point on the reference genome, and then compute the number of mismatches in the corresponding alignment. These computations are dynamically terminated so that only 'unique' alignments are reported, according to the following rule: if an alignment *A* has only *x* mismatches, and if there is no alternative alignment having $\leq x + 2$ mismatches, then we call *A* unique. To minimize the risk of amplification bias, only one read was kept if multiple reads aligned to the same start point.

For each ChIP (or control) experiment, we next estimate the number of end-sequenced ChIP fragments that overlap any given nucleotide position in the reference genome (here, at 25-bp resolution). For each position, we count the number of aligned reads that are oriented towards it and closer than the average length of a library fragment (~300 bp).

To identify the portion of the mouse genome that can be interrogated with SMS reads of a given length (k) and alignment stringency, we aligned every k-mer that occurs in the reference sequence (mm8) using the same pipeline as for SMS reads. Nucleotide positions in the reference genome where less than 50% of the 200 flanking k-mers on each side had 'unique' alignments were masked as repetitive and disregarded from further analysis (<28% of the genome). Although we analysed reads spanning 27–36 bp, all data were conservatively masked at $k = 27$.

We identified genomic intervals enriched with a specific chromatin mark from the mean fragment count in 1-kb sliding windows. To account for varying read numbers and lengths, we generated sample-specific expected distributions of fragment counts under the null hypothesis of no enrichment by moving each aligned read to a randomly chosen 'unique' position on the same chromosome. Nominal *P*-values for enrichment at a particular position were obtained by comparison to a randomized version of the same data set (due to the large number of reads, multiple randomizations gave identical results). Genome-wide maps of enriched sites were created by identification of windows where the nominal *P*-value fell below 10^{-5}, and merging any enriched windows that were less than 1-kb apart into continuous intervals. To improve sensitivity to the more diffuse enrichment observed from H3K9me3 and H4K20me3 near repetitive regions and from H3K36me3 across large transcripts, we also developed a Hidden Markov Model (HMM) to segment the reference genome into 'enriched' and 'unenriched' intervals (R.P.K., manuscript in preparation). The observed fragment densities were divided into four categories, in a sample-dependent manner ('masked', 'sub-threshold', 'near-threshold' and 'above threshold'). Emission and transition probabilities were fitted using supervised learning on limited intervals (~10 Mb total) chosen to reflect diverse chromatin landscapes, and the resultant models were applied genome wide.

Validation of ChIP-Seq by comparison to ChIP-chip and real-time PCR. ChIP-Seq data for H3K4me3 and H3K27me3 in ES cells were compared to published ChIP-chip profiles across ~2% of the mouse genome[9]. Significantly enriched sites in the ChIP-chip data were defined using a previously validated *P*-value threshold of 10^{-4}, and compared to the ChIP-Seq sites. In addition, a set of 50 PCR primer pairs (Supplementary Table 2) was designed to amplify 100–140-bp fragments from genomic regions showing a wide range of signals for H3K4me3 and H3K27me3 by ChIP-Seq. Real-time PCR was carried out using Quantitect SYBR green PCR mix (Qiagen) on a 7000 ABI detection system, using 0.25 ng ChIP or WCE DNA as template. Fold enrichments reflect two independent ChIP assays, each evaluated in duplicate by real-time PCR.

Promoter classification and definition of gene and transcript intervals. The analysed promoters were based on transcription start sites inferred from full-length mouse RefSeqs (downloaded from the UCSC Genome Brower 2 April 2007). Promoters containing a 500-bp interval within −0.5 kb to +2 kb with a (G+C)-fraction ≥0.55 and a CpG observed to expected ratio (O/E) ≥0.6 were classified as HCPs. Promoters containing no 500-bp interval with CpG O/E ≥0.4 were classified as LCPs. The remainder were classified as ICPs. The chromatin states of promoters were determined by overlap with cell-type-specific H3K4me3 and H3K27me3 intervals. For comparison with expression levels, the chromatin states of genes with more than one known promoter were classified according to the most 'active' mark (that is, a gene with an H3K4me3 marked promoter and a bivalent promoter would be classified as 'H3K4me3'). Correlation between H3K4me3 enrichment and expression levels was calculated from the mean fragment density over each promoter from −0.5 kb to +1 kb. Correlation between H3K36me3 and expression levels was calculated from the mean fragment density over each RefSeq transcript.

Expression data. RNA expression data for ES cells, NPCs and MEFs were generated from polyA RNA using GeneChip Mouse Genome 430 2.0 Arrays (Affymetrix). Expression data for adult tissues were downloaded from the Novartis Gene Expression Atlas at http://symatlas.gnf.org. Pre-processing, normalization (GC-RMA) and hierarchical clustering (Pearson, log-transformed, row-centred values) were performed using GenePattern (http://www.broad.mit.edu/cancer/software/).

Analysis of repetitive elements. Chromatin state at repetitive elements was evaluated by aligning SMS reads directly to a library of repetitive element consensus sequences (http://www.girinst.org). The proportion of reads aligning to each class was calculated for H3K9me3 and H4K20me3, and enrichment determined by comparison to WCE and pan-H3. We also applied an orthogonal approach based on HMM intervals of H3K9me3 in unique sequences (see above). For each repetitive element type or class, we calculated the number of occurrences within 1 kb of a unique H3K9me3 site, controlling against a set of randomly placed sites of the same length distribution.

Allele-specific histone methylation. SNPs between the 129S1/SvlmJ (used as proxy for 129SvJae) and *M. m. castaneus* mouse strains were obtained from Perlegen at http://mouse.perlegen.com. Allele-specific bias was evaluated by a binomial test of the null hypothesis that ChIP fragments were drawn uniformly from both alleles. (H3K4me3 and H3K9me3 reads were pooled before the test, see Supplementary Table 7.) We note that the 129SvJae strain is closer to the C57BL/6-derived reference genome, and this may cause a slight bias towards assigning aligned reads to this strain. To minimize this bias, aligned reads were kept for analysis if no alternative alignment had the same number of mismatches to the reference sequence.

ARTICLES

Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer

Manabu Soda[1,2], Young Lim Choi[1], Munehiro Enomoto[1,2], Shuji Takada[1], Yoshihiro Yamashita[1], Shunpei Ishikawa[5], Shin-ichiro Fujiwara[1], Hideki Watanabe[1], Kentaro Kurashina[1], Hisashi Hatanaka[1], Masashi Bando[2], Shoji Ohno[2], Yuichi Ishikawa[6], Hiroyuki Aburatani[5,7], Toshiro Niki[3], Yasunori Sohara[4], Yukihiko Sugiyama[2] & Hiroyuki Mano[1,7]

Improvement in the clinical outcome of lung cancer is likely to be achieved by identification of the molecular events that underlie its pathogenesis. Here we show that a small inversion within chromosome 2p results in the formation of a fusion gene comprising portions of the echinoderm microtubule-associated protein-like 4 (*EML4*) gene and the anaplastic lymphoma kinase (*ALK*) gene in non-small-cell lung cancer (NSCLC) cells. Mouse 3T3 fibroblasts forced to express this human fusion tyrosine kinase generated transformed foci in culture and subcutaneous tumours in nude mice. The *EML4–ALK* fusion transcript was detected in 6.7% (5 out of 75) of NSCLC patients examined; these individuals were distinct from those harbouring mutations in the epidermal growth factor receptor gene. Our data demonstrate that a subset of NSCLC patients may express a transforming fusion kinase that is a promising candidate for a therapeutic target as well as for a diagnostic molecular marker in NSCLC.

Lung cancer remains the leading cause of cancer deaths in western countries[1]. Patients with NSCLC, which accounts for ~80% of lung cancer cases, are often diagnosed at advanced stages of the disease. Given that conventional chemotherapeutic regimens only marginally improve the outcome of such individuals, their median survival time is less than one year after diagnosis (ref. 2). A subset of NSCLCs was recently shown to harbour activating mutations in the epidermal growth factor receptor gene (*EGFR*)[3,4]; such cancers are responsive to gefitinib, a specific inhibitor of the tyrosine kinase activity of EGFR. The efficacy of targeting key 'growth drivers' in cancer treatment is further exemplified by chronic myeloid leukaemia, for which another tyrosine kinase inhibitor, STI571, is highly effective in reducing the number of cancer cells[5]. However, *EGFR* mutations are associated preferentially with NSCLC of non-smokers and Asians[4,6]. Few oncogenes have thus been identified for NSCLC in individuals with a smoking habit, who constitute most cases of the disease.

Retrovirus-mediated complementary DNA expression systems allow expression of the encoded proteins in most of the targeted cells. Through modification of the method used in ref. 7, we have achieved reliable amplification of cDNAs from small quantities of clinical specimens as well as the generation of retroviral libraries for expression of these cDNAs[8–10]. Application of such a cDNA expression library prepared from an NSCLC specimen to a focus formation assay with mouse 3T3 fibroblasts has now led to the identification of a fusion oncogene.

Identification of *EML4–ALK*

To isolate novel transforming genes in NSCLC, we generated a retroviral cDNA expression library from a lung adenocarcinoma specimen surgically resected from a 62-yr-old man with a history of smoking (patient 33). In construction of the library, we used the SMART method (Clontech) for preferential amplification of full-length cDNAs from limited amounts of clinical specimens[7]; this resulted in the production of $>1.4 \times 10^6$ independent plasmid clones. Infection of mouse 3T3 fibroblasts with the recombinant retroviruses that were based on these plasmids led to the formation of many transformed foci, from which insert cDNAs were recovered with the polymerase chain reaction (PCR).

One of the amplified cDNAs comprised 3,926 base pairs (bp) and contained an open reading frame for a protein of 1,059 amino acids (Fig. 1a and Supplementary Fig. 1). The amino-terminal portion (residues 1–496) of the predicted protein is identical to that of human EML4 (GenBank accession number NM_019063), whereas the carboxy-terminal portion (residues 497–1059) is identical to the intracellular domain (residues 1058–1620 of the wild-type protein) of human ALK (GenBank accession number AB209477), suggesting that the cDNA is derived from a fusion product of *EML4* and *ALK* (Fig. 1a, b). EML4 belongs to the family of echinoderm microtubule-associated protein-like proteins[11] and is composed of an N-terminal basic region (isoelectric point, 10.2), a hydrophobic echinoderm microtubule-associated protein-like protein (HELP) domain[12] and WD repeats[13] (Fig. 1b). In the predicted fusion protein, the N-terminal half of EML4 encompassing the basic region, the HELP domain and a portion of the WD-repeat region is fused to the intracellular juxtamembrane region of ALK.

ALK was first identified as a fusion partner of nucleophosmin (NPM) in anaplastic large-cell lymphoma with a t(2;5) chromosome rearrangement[14,15]. Other chromosome translocations involving the *ALK* locus were subsequently identified in the same lymphoma subtype as well as in inflammatory myofibroblastic tumours[16]. The fusion point of ALK is conserved among most of these chimaeric tyrosine kinases, including EML4–ALK (resulting in fusion of the

[1]Division of Functional Genomics, [2]Division of Pulmonary Medicine, [3]Department of Pathology, and [4]Division of General Thoracic Surgery, Jichi Medical University, Tochigi 329-0498, Japan. [5]Research Center for Advanced Science and Technology, University of Tokyo, Tokyo 153-8904, Japan. [6]Department of Pathology, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. [7]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Saitama 332-0012, Japan.

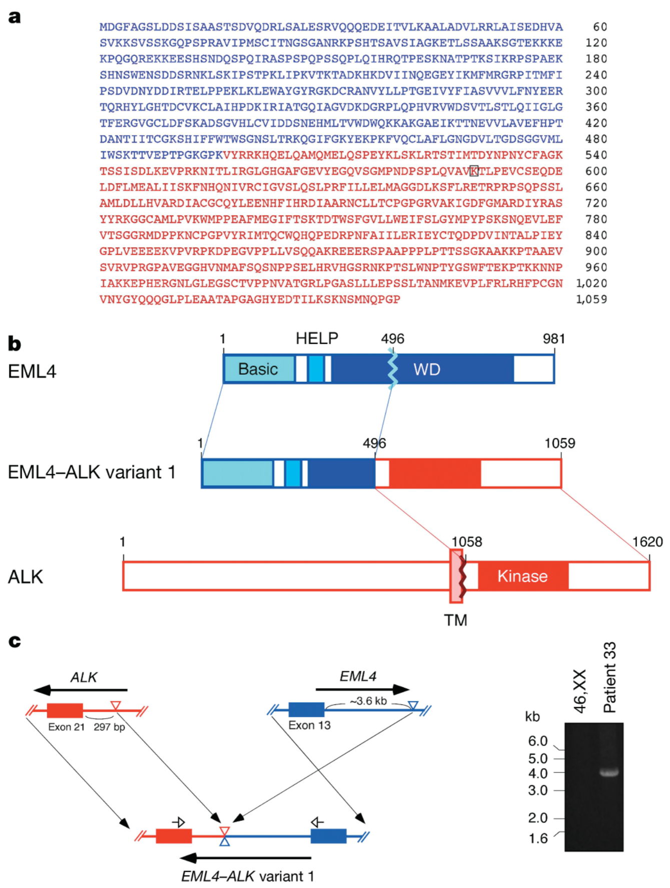


Figure 1 | Gene fusion between *EML4* and *ALK*. **a**, Amino acid sequence of the EML4–ALK protein (variant 1). Residues corresponding to EML4 or to ALK are shown in blue or red, respectively. Lys 589 in the ATP-binding site is boxed. **b**, Fusion of the N-terminal portion of EML4 (comprising the basic region, the HELP domain and part of the WD-repeat region) to the intracellular region of ALK (containing the tyrosine kinase domain). TM, transmembrane domain. **c**, Both the *ALK* gene and the *EML4* gene map to chromosome 2p, but have opposite orientations. In the NSCLC patient 33, *EML4* is disrupted at a position ~3.6 kb downstream of exon 13 and is ligated to a position 297 bp upstream of exon 21 of *ALK*, giving rise to the *EML4–ALK* (variant 1) fusion gene (left panel). Filled and open horizontal arrows indicate the direction of transcription and the positions of the Fusion-genome primers, respectively (Supplementary Fig. 1). PCR with these Fusion-genome primers and genomic DNA of patient 33 generated a single product of ~4 kb (right panel); this product was not detected with control DNA of a healthy female (46,XX).

entire intracellular kinase domain of ALK to the corresponding partner), and the kinase activity of NPM–ALK was shown to be essential for the proliferation of lymphoma cells positive for this construct[17].

Given that *EML4* and *ALK* each map to the short arm of chromosome 2 (2p21 and 2p23, respectively, separated by a distance of ~12 megabases, Mb) but have opposite orientations, either gene might have been inverted to generate the *EML4–ALK* fusion gene (Fig. 1c). To address this issue directly, we amplified the genomic fusion point between *EML4* and *ALK* using genomic DNA of patient 33 as the template. This approach led to the identification of a ~4-kilobase (kb) product (Fig. 1c). Nucleotide sequencing of this genomic fragment revealed that intron 13 of *EML4* is disrupted at a point ~3.6 kb downstream of exon 13 and is inverted to connect to a position 297 bp upstream of exon 21 of *ALK* (Fig. 1c and Supplementary Data), yielding *EML4–ALK* variant 1 (the structure of variant 2 is addressed below). To determine whether the chromosome rearrangement in this specimen is a simple inversion within 2p, we attempted to detect the other connection point between *EML4* and *ALK* in chromosome 2 by PCR amplification of the *ALK–EML4* cDNA or gene. However, neither of the corresponding PCR products was obtained (data not shown). It thus remains undetermined whether the NSCLC cells of patient 33 harbour a simple inv(2)(p21p23) or whether they contain complex chromosome translocations involving 2p.

Transforming activity of EML4–ALK

To confirm the transforming potential of EML4–ALK, we generated expression plasmids for wild-type EML4, wild-type ALK, EML4–ALK, EML4–ALK(K589M) (in which Lys 589 in the ATP-binding site of the kinase domain is replaced with Met), NPM–ALK and v-Ras, and introduced them individually into mouse 3T3 fibroblasts. Transformed foci were readily identified for the cells expressing EML4–ALK, NPM–ALK or v-Ras, but not for those expressing EML4, ALK or EML4–ALK(K589M) (Fig. 2a). Subcutaneous injection of the transfected 3T3 cells into nude mice also revealed that only those expressing EML4–ALK, NPM–ALK or v-Ras formed tumours (Fig. 2a). These data thus showed that EML4–ALK possesses transforming activity that is dependent on its catalytic activity. We also found that fusion to EML4 results in redistribution of the kinase domain of ALK from the cell membrane to the cytoplasm (Supplementary Fig. 2), as revealed by monitoring the fluorescence of the corresponding proteins tagged with enhanced green fluorescent protein.

To identify the domains of EML4 required for the transforming activity of the EML4–ALK fusion protein, we generated expression plasmids for EML4–ALK with internal deletions of the basic domain (ΔBasic, lacking residues 31–140), of the HELP domain (ΔHELP, lacking residues 220–296) or of the WD repeats (ΔWD, lacking residues 305–475) (Fig. 2b). Injection of 3T3 cells expressing the deletion constructs into nude mice revealed that deletion of the WD repeats allowed tumour formation at all injection sites, but that the tumours were smaller than those formed by cells expressing full-length EML4–ALK (Fig. 2b). Tumours were even smaller for cells expressing ΔHELP and were undetectable for those expressing ΔBasic. All domains of EML4 thus seem to contribute to the oncogenic potential of EML4–ALK, with the basic domain being the most important.

Gene fusion often results in activation of tyrosine kinases through oligomerization mediated by the fusion partner. Although little is known regarding the dimerization potential of EML4, our results (Fig. 2b) suggested that the basic domain derived from EML4 may mediate EML4–ALK dimerization. To examine this possibility, we transfected HEK 293 cells with expression vectors for both Myc-epitope-tagged EML4–ALK and Flag-epitope-tagged full-length, ΔBasic, ΔHELP or ΔWD forms of EML4–ALK. Immunoprecipitation of cell lysates with antibodies to Myc and probing of the resulting precipitates with antibodies to Flag revealed that Myc-epitope-tagged EML4–ALK was associated with substantial amounts of each of the Flag-tagged EML4–ALK constructs with the exception of ΔBasic (Fig. 2c). Immunoblot analysis of immunoprecipitates prepared from the same cell lysates with antibodies to Flag confirmed that the various Flag-tagged EML4–ALK constructs were expressed at similar levels. Similar results were obtained in a reciprocal experiment in which anti-Flag immunoprecipitates were probed with anti-Myc (Supplementary Fig. 3). These results thus indicated that the basic domain indeed has an important role in dimerization of EML4–ALK.

To examine directly whether internal deletions affect the enzymatic activity of EML4–ALK, we expressed Flag-tagged full-length or truncated forms of the fusion protein in HEK 293 cells and prepared immunoprecipitates from cell lysates with anti-Flag. The resulting precipitates were then subjected to an *in vitro* kinase assay with the synthetic YFF peptide[18], which is based on the activation loop of the catalytic domain of ALK. Deletion of the basic domain resulted in a marked decrease (~84%) in the catalytic activity of EML4–ALK (Fig. 2d). This low level of kinase activity of the ΔBasic mutant was consistent with its residual ability to dimerize with the full-length

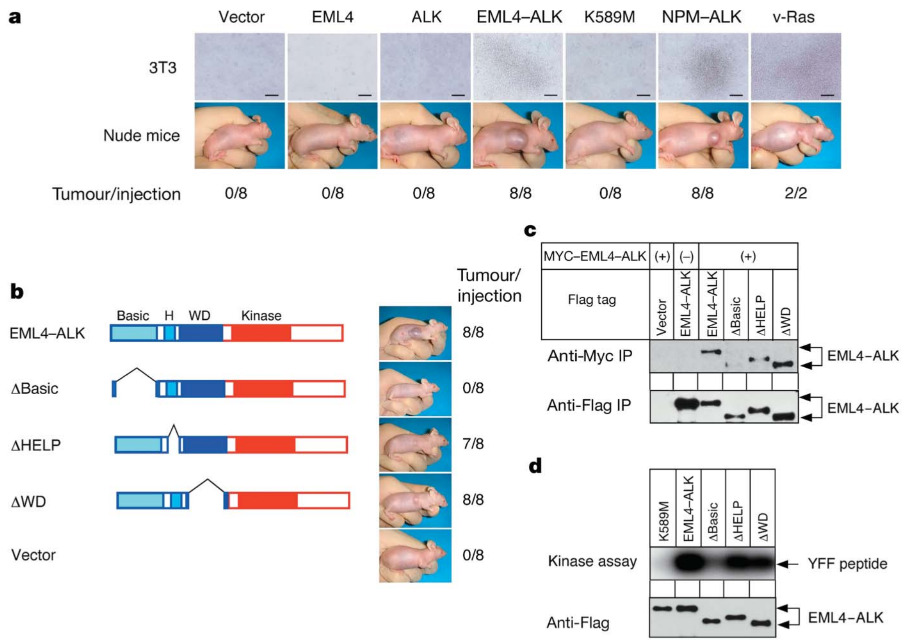


Figure 2 | Transforming activity of EML4–ALK variant 1. a, Expression vectors for EML4, ALK, EML4–ALK, EML4–ALK(K589M), NPM–ALK and v-Ras (or the corresponding empty vector) were introduced individually into 3T3 cells, and the cells were photographed after three weeks of culture (upper panels). Scale bars, 100 μm. The same set of transfected cells was also injected subcutaneously into nude mice, and tumour formation was examined after 20 days (lower panels). The number of tumours formed after eight or two injections is indicated. **b**, Schematic representations of EML4–ALK and its deletion mutants are shown on the left. Tumour formation in nude mice was examined as in **a** for 3T3 cells transfected with expression vectors for the indicated forms of EML4–ALK (right). H, HELP domain. **c**, Expression vectors for Flag-tagged EML4–ALK or its deletion mutants were introduced into HEK 293 cells with (+) or without (–) a vector for Myc-epitope-tagged EML4–ALK. Cell lysates were subjected to immunoprecipitation (IP) with antibodies to Myc or to Flag, and the resulting precipitates were subjected to immunoblot analysis with anti-Flag. The positions of EML4–ALK and its mutants are shown on the right. **d**, Expression vectors for Flag-tagged EML4–ALK, EML4–ALK(K589M) or its deletion mutants were introduced into HEK 293 cells. Immunoprecipitates prepared from cell lysates with anti-Flag were subjected to an *in vitro* kinase assay with the synthetic YFF peptide (upper panel) or to immunoblot analysis with anti-Flag (lower panel).

protein (Fig. 2c and Supplementary Fig. 3). Deletion of either the HELP or WD domain reduced the kinase activity of EML4–ALK by ~50% (Fig. 2d), whereas the transforming activity of the ΔHELP mutant was reproducibly lower than that of ΔWD in both the focus formation (data not shown) and tumorigenicity (Fig. 2b) assays. The molecular basis of this discrepancy between kinase and transforming activities remains to be determined.

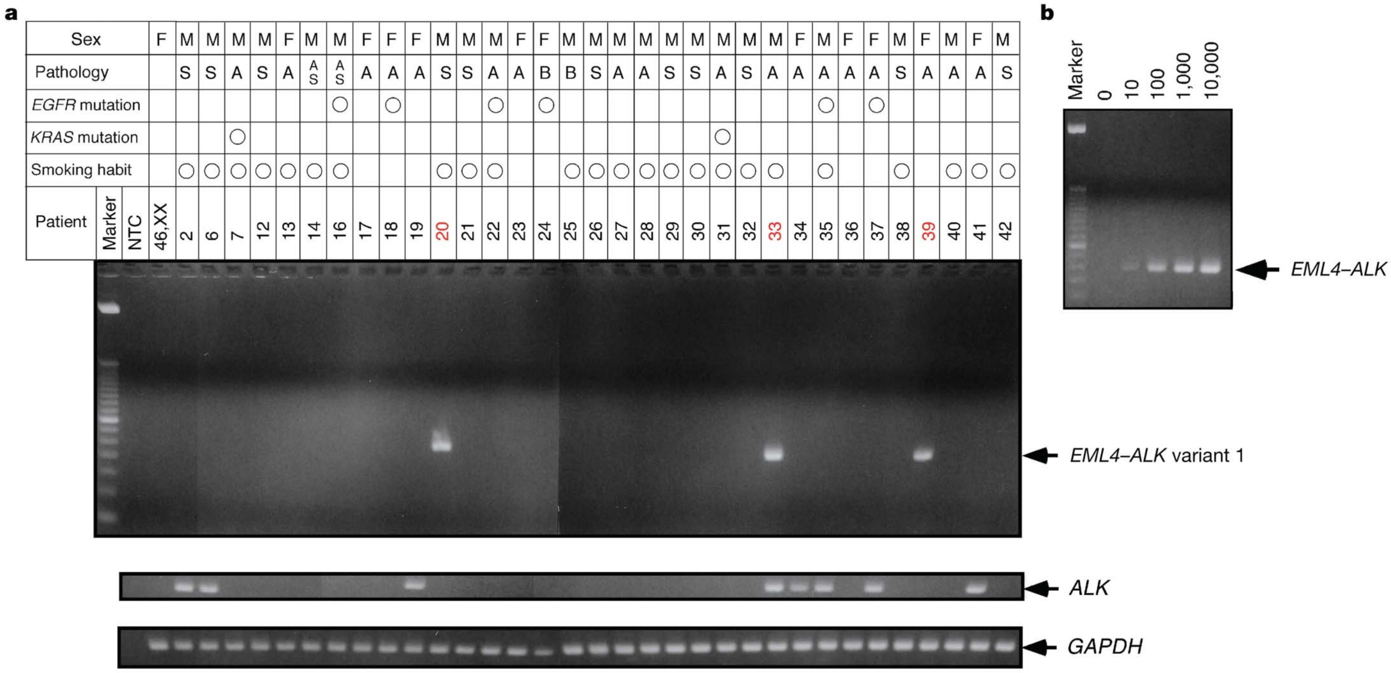
Detection of *EML4–ALK* in clinical specimens

We next evaluated the frequency of *EML4–ALK* gene fusion in NSCLC. A consecutive panel of NSCLC specimens ($n = 33$) obtained in one hospital was examined for the presence of *EML4–ALK* messenger RNA, wild-type *ALK* mRNA and mutations within the *EGFR* and v-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue (*KRAS*) genes. The *EML4–ALK* fusion mRNA was readily detected by PCR with reverse transcription (RT–PCR) analysis as a 247-bp product in three patients (9.1%), including the patient who served as the source for the retroviral library (Fig. 3a). Sequencing of the PCR products amplified from each of these three patients confirmed the presence of *EML4–ALK* variant 1 cDNA (Supplementary Data); patients 20 and 39 had squamous cell carcinoma and adenocarcinoma of the lung, respectively.

We also amplified the genomic fragments corresponding to exons 18, 19 and 21 of *EGFR* and determined their nucleotide sequences in all 33 patients. This revealed the presence of *EGFR* mutations—all of which were deletions or nucleotide substitutions within exon 19—in six individuals (18.2%) (Supplementary Table 1). Notably, the patient population harbouring *EGFR* mutations did not overlap with that harbouring the *EML4–ALK* fusion gene, showing that *EML4–ALK*-positive cancer is a novel subclass within NSCLC. A *KRAS* mutation (Val 12 to Cys 12 substitution) was detected in two individuals, neither of whom harboured *EGFR* mutations or the *EML4–ALK* fusion gene. Wild-type *ALK* mRNA was detected in 8 of the 33 specimens (24%) of this cohort. A moderate level of *ALK* mRNA in lung cancer specimens has also been detected by serial analysis of gene expression studies (http://cgap.nci.nih.gov/SAGE/AnatomicViewer), although it is not clear whether such profiling of the 3′ end of mRNAs actually detected mRNAs for wild-type *ALK*, *EML4–ALK* or even other *ALK* fusion genes.

To determine whether fusion of *EML4* to *ALK* is specific to NSCLC, we used RT–PCR to attempt to detect the fusion mRNA in cancer specimens from 39 patients with acute myeloid leukaemia, 69 patients with non-Hodgkin's lymphoma, 93 patients with gastric carcinoma and 60 patients with colorectal carcinoma. However, none of these 261 specimens yielded the *EML4–ALK* cDNA (data not shown), indicating that *EML4–ALK* has a high level of specificity to NSCLC.

Further screening for the *EML4–ALK* fusion cDNA with the same RT–PCR primer set in a different cohort of NSCLC patients ($n = 42$) resulted in the identification of a larger PCR product (~1 kb) in another two individuals with lung adenocarcinoma. Nucleotide sequencing of these PCR products revealed that exon 20 of *EML4* was fused to exon 21 of *ALK* (Supplementary Fig. 4 and Supplementary Data), indicative of diversity in the breakpoint region within *EML4*. PCR analysis of genomic DNA from one of these two patients further revealed the breakpoints in *EML4* and *ALK* as well as the formation of both *EML4–ALK* and *ALK–EML4* fusion genes, thus demonstrating the presence of an inv(2)(p21p23) rearrangement in this individual (Supplementary Fig. 4 and Supplementary Data). We here refer to the initially identified *EML4–ALK* gene, in which intron 13 of *EML4* is fused to intron 20 of *ALK*, as variant 1, and to the *EML4–ALK* gene, in which intron 20 of *EML4* is fused to intron 20 of



	Marker	NTC	46,XX	2	6	7	12	13	14	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Sex			F	M	M	M	M	F	M	M	F	F	F	M	M	M	F	F	M	M	M	M	M	M	M	M	M	F	M	F	F	M	F	M	F	M
Pathology				S	S	A	S	A	A/S	A/S	A	A	A	S	S	A	A	B	B	S	A	A	S	S	A	S	A	A	A	A	A	S	A	A	A	S
EGFR mutation										○		○				○		○											○		○					
KRAS mutation						○																			○											
Smoking habit				○	○	○	○	○	○	○				○	○	○			○	○	○	○	○	○	○	○	○		○			○		○	○	○

Figure 3 | Screening of NSCLC specimens for *EML4–ALK* variant 1 mRNA. **a**, A consecutive panel of NSCLC specimens ($n = 33$) was subjected to RT–PCR with the Fusion-RT primers (Supplementary Fig. 1). Patients positive for the *EML4–ALK* (variant 1) PCR product are shown in red. Peripheral blood mononuclear cells from a healthy female (46,XX) were also examined as a negative control. RT–PCR for wild-type *ALK* mRNA (with a primer set corresponding to the extracellular domain of ALK) and for glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA is also shown. Patient characteristics (sex, pathological classification of NSCLC, the presence of *EGFR* or *KRAS* mutations, and smoking habit) are indicated at the top. A, adenocarcinoma; AS, adenosquamous carcinoma; B, bronchiolo-alveolar carcinoma; F, female; M, male; NTC, no-template control; S, squamous carcinoma. Marker, 50-bp DNA ladder. **b**, Sputum (1 ml) was mixed with 0, 10, 100, 1,000 or 10,000 BA/F3 cells expressing *EML4–ALK* (variant 1) and was then subjected to RT–PCR with the Fusion-RT primer set for detection of *EML4–ALK* mRNA. Marker, 50-bp DNA ladder.

ALK, as variant 2. Multiple variants of the TRK-fused gene (*TFG*)–*ALK* fusion gene associated with anaplastic large-cell lymphoma have also been identified[19].

Given the head-to-head orientation of *EML4* and *ALK* on chromosome 2, RT–PCR with the primers used in our study would not be expected to yield specific products in normal tissues or in any cancers that do not harbour the fusion gene. RT–PCR for *EML4–ALK* mRNA may thus provide a highly sensitive means for detection of lung cancer (with the corresponding chromosomal rearrangement). Although cytological examination of sputum is a reliable method for detection of lung cancer, it is usually only effective at advanced stages. Detection of *EGFR* mutations in sputum is also problematic because NSCLC cells often constitute only a small fraction of cells within the specimen. In contrast, RT–PCR would be expected to detect the presence of only a few cells harbouring the *EML4–ALK* fusion gene among the tens of thousands of non-cancerous cells in sputum. To examine this issue, we established mouse BA/F3 cells[20] that express Flag-tagged EML4–ALK, mixed various numbers of these cells with control sputum, and subjected the mixtures to RT–PCR for detection of *EML4–ALK* mRNA. The fusion mRNA was detected in sputum containing as few as ten BA/F3 cells per ml (Fig. 3b). Early diagnosis of NSCLC (of the *EML4–ALK*$^+$ subtype) by RT–PCR may thus be realistic, as is clinical detection of *Mycobacterium tuberculosis* in sputum by PCR[21]. It is also possible that *EML4–ALK* mRNA would be detected by RT–PCR using pleural effusion, bronchoalveolar lavage, lung biopsy or peripheral blood specimens of patients with lung cancer.

EML4–ALK as a potential therapeutic target

Several small compounds have recently been shown to inhibit the kinase activity of ALK and to suppress the growth of cells expressing NPM–ALK[17,22,23]. To investigate whether use of such ALK inhibitors might be an effective treatment for EML4–ALK$^+$ NSCLC, we expressed Flag-tagged ALK, EML4–ALK or EML4–ALK(K589M) in BA/F3 cells, which are dependent on interleukin-3 (IL-3) for growth. Whereas all transfected cells grew exponentially in the presence of IL-3, only those expressing EML4–ALK proliferated at a similar rate in the absence of IL-3 (Fig. 4a), again confirming the kinase-dependent oncogenic activity of EML4–ALK.

In the presence of IL-3, BA/F3 cells proliferate in a manner dependent on the activity of the tyrosine kinase Janus kinase 2 (JAK2, ref. 24). Addition of a chemical inhibitor (WHI-P154)[22] of ALK to the culture medium affected the IL-3-dependent growth of BA/F3 cells only slightly at concentrations $\geq 5\,\mu M$ (Fig. 4b). Given that WHI-P154 was originally developed as a specific inhibitor of JAK3, it might be expected to show a weak cross-reactivity with JAK2, possibly accounting for the small effect on the JAK2-dependent growth of BA/F3 cells. In contrast, WHI-P154 markedly inhibited the growth of BA/F3 cells expressing EML4–ALK, which do not require IL-3 for proliferation (Fig. 4b). At a concentration of $10\,\mu M$, WHI-P154 rapidly induced the death of these cells in the absence of IL-3. Consistent with these observations, immunoblot analysis revealed that WHI-P154 inhibited the tyrosine phosphorylation of EML4–ALK (on the residue corresponding to Tyr 1604 of wild-type ALK) in a concentration-dependent manner in the transfected BA/F3 cells (Fig. 4c).

Discussion

Using retrovirus-mediated expression screening, we have identified an oncogene, *EML4–ALK*, in a specimen of NSCLC. The 75 NSCLC patients examined in the present study (5 of whom were positive for *EML4–ALK*) were all Japanese. Given that the association of *EGFR* mutations with lung cancer is most prominent in Asian populations[5], it will be important to examine the association of *EML4–ALK* with lung cancer in other ethnic groups. Our data obtained with WHI-P154 suggest that inhibition of the tyrosine kinase activity of

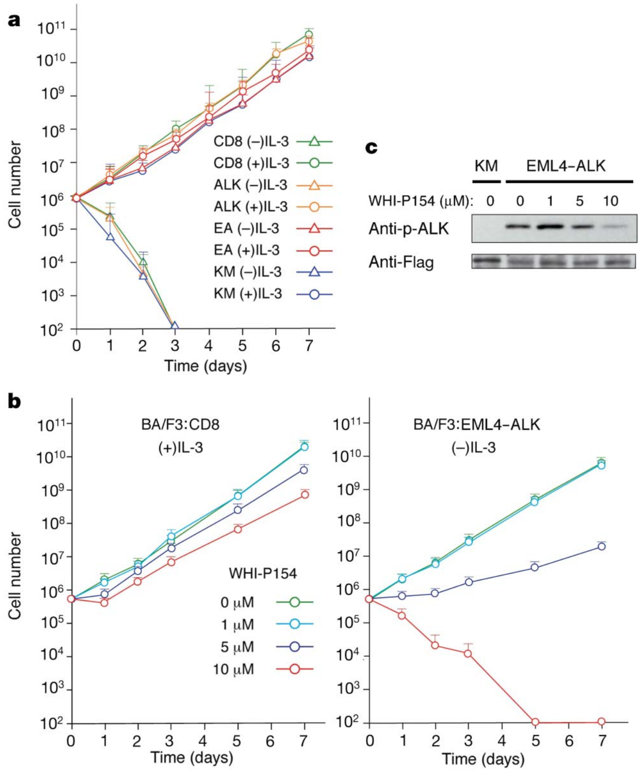


Figure 4 | Inhibition of the growth of BA/F3 cells expressing EML4–ALK variant 1 by a chemical inhibitor of ALK. a, Mouse BA/F3 cells expressing CD8 either alone or together with wild-type ALK, EML4–ALK (EA) or EML4–ALK(K589M) (KM) were cultured in the absence (–) or presence (+) of IL-3 (1 ng ml^{-1}). Cell number was determined at the indicated times. Data are means plus s.d. of values from three separate experiments. **b**, BA/F3 cells expressing CD8 alone were cultured with IL-3 and 0, 1, 5 or 10 μM WHI-P154 (left panel), or those expressing CD8 and EML4–ALK were incubated without IL-3 but with 0, 1, 5 or 10 μM WHI-P154 (right panel). Cell number was determined at the indicated times. Data are means plus s.d. of values from three separate experiments. **c**, BA/F3 cells expressing Flag-tagged EML4–ALK or EML4–ALK(K589M) were incubated with the indicated concentrations of WHI-P154 for 3.5 h, after which total cell lysates (25 μg of protein per lane) were subjected to immunoblot analysis with antibodies to tyrosine-phosphorylated ALK (p-ALK) or to Flag.

EML4–ALK may induce cell death in tumours expressing this fusion protein. Given the lack of apparent phenotypes in *Alk* knockout mice[25], suppression of EML4–ALK function with ALK inhibitors might be expected to be free of severe side effects in NSCLC patients. Furthermore, given that the population of patients who harbour *EGFR* mutations is distinct from that which harbours the *EML4–ALK* fusion gene, ALK inhibitors may provide a means to control NSCLC in the latter population of patients, for whom effective treatments are rarely available.

METHODS SUMMARY

A recombinant retroviral cDNA expression library was constructed as described previously[7–10] from a lung cancer specimen, and was used to infect mouse 3T3 fibroblasts. Transformed foci isolated from the cells after 2 weeks of culture were subjected to extraction of genomic DNA and amplification of retroviral insert cDNAs by PCR. The *EML4–ALK* fusion cDNA was detected by RT–PCR analysis of total RNA from clinical specimens with the Fusion-RT-S (5′-GTGCAGTGTTTAGCATTCTTGGGG-3′) and Fusion-RT-AS (5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′) primers. The *EML4–ALK* variant 1 gene was detected by PCR with genomic DNA of clinical specimens and the Fusion-genome-S (5′-CCACACCTGGGAAAGGACCTAAAG-3′) and Fusion-genome-AS (5′-AGCTTGCTCAGCTTGTACTCAGGG-3′) primers. Expression plasmids for Flag-epitope-tagged ALK, EML4 and EML4–ALK were generated with the retroviral vector pMXS[26]. The kinase-inactive mutant (K589M) of EML4–ALK was constructed by site-directed mutagenesis. Internal deletion mutants of EML4–ALK were also generated by mutagenesis. The expression plasmids were introduced into 3T3 cells by the calcium phosphate method, and the cells were then either cultured for 21 days or injected subcutaneously into nude mice. For analysis of EML4–ALK dimerization, expression vectors for Flag- or Myc-epitope-tagged EML4–ALK or its mutants were introduced into HEK 293 cells, cell lysates were subjected to immunoprecipitation with anti-Flag or anti-Myc, and the resulting precipitates were subjected to immunoblot analysis with the same antibodies. Anti-Flag immunoprecipitates were also subjected to an *in vitro* kinase assay with the synthetic YFF peptide[18]. For BA/F3 experiments, cDNAs for ALK, EML4–ALK or EML4–ALK(K589M) were inserted into the plasmid pMX-iresCD8 (ref. 27) to confer simultaneous expression of the protein of interest and mouse CD8. BA/F3 cells were infected with recombinant retroviruses generated from each plasmid, and the resulting CD8^{+} cells were purified and incubated with various concentrations of WHI-P154 (EMD Biosciences).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 February; accepted 17 May 2007.
Published online 11 July 2007.

1. Jemal, A. *et al.* Cancer statistics, 2006. *CA Cancer J. Clin.* **56,** 106–130 (2006).
2. Schiller, J. H. *et al.* Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N. Engl. J. Med.* **346,** 92–98 (2002).
3. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350,** 2129–2139 (2004).
4. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).
5. Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344,** 1031–1037 (2001).
6. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA* **101,** 13306–13311 (2004).
7. Yoshizuka, N. *et al.* An alternative transcript derived from the *Trio* locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential. *J. Biol. Chem.* **279,** 43998–44004 (2004).
8. Hatanaka, H. *et al.* Transforming activity of purinergic receptor P2Y, G-protein coupled, 2 revealed by retroviral expression screening. *Biochem. Biophys. Res. Commun.* **356,** 723–726 (2007).
9. Choi, Y. L. *et al.* Identification of a constitutively active mutant of JAK3 by retroviral expression screening. *Leuk. Res.* **31,** 203–209 (2007).
10. Fujiwara, S. *et al.* Transforming activity of the lymphotoxin-β receptor revealed by expression screening. *Biochem. Biophys. Res. Commun.* **338,** 1256–1262 (2005).
11. Pollmann, M. *et al.* Human EML4, a novel member of the EMAP family, is essential for microtubule formation. *Exp. Cell Res.* **312,** 3241–3251 (2006).
12. Eichenmuller, B., Everley, P., Palange, J., Lepley, D. & Suprenant, K. A. The human EMAP-like protein-70 (ELP70) is a microtubule destabilizer that localizes to the mitotic apparatus. *J. Biol. Chem.* **277,** 1301–1309 (2002).
13. Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* **24,** 181–185 (1999).
14. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263,** 1281–1284 (1994).
15. Shiota, M. *et al.* Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* **9,** 1567–1574 (1994).
16. Pulford, K., Morris, S. W. & Turturro, F. Anaplastic lymphoma kinase proteins in growth control and cancer. *J. Cell. Physiol.* **199,** 330–358 (2004).
17. Galkin, A. V. *et al.* Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc. Natl Acad. Sci. USA* **104,** 270–275 (2007).
18. Donella-Deana, A. *et al.* Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* **44,** 8533–8542 (2005).
19. Hernandez, L. *et al.* Diversity of genomic breakpoints in *TFG-ALK* translocations in anaplastic large cell lymphomas: identification of a new TFG-ALK$_{XL}$ chimeric gene with transforming activity. *Am. J. Pathol.* **160,** 1487–1494 (2002).
20. Palacious, R. & Steinmetz, M. IL-3 dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ-line configuration, and generate B lymphocytes *in vivo*. *Cell* **41,** 727–734 (1985).
21. Kaul, K. L. Molecular detection of *Mycobacterium tuberculosis*: impact on patient care. *Clin. Chem.* **47,** 1553–1558 (2001).
22. Marzec, M. *et al.* Inhibition of ALK enzymatic activity in T-cell lymphoma cells induces apoptosis and suppresses proliferation and STAT3 phosphorylation independently of Jak3. *Lab. Invest.* **85,** 1544–1554 (2005).
23. Li, R. *et al.* Design and synthesis of 5-aryl-pyridone-carboxamides as inhibitors of anaplastic lymphoma kinase. *J. Med. Chem.* **49,** 1006–1015 (2006).
24. Watanabe, S., Itoh, T. & Arai, K. JAK2 is essential for activation of c-*fos* and c-*myc* promoters and cell proliferation through the human granulocyte-macrophage colony-stimulating factor receptor in BA/F3 cells. *J. Biol. Chem.* **271,** 12681–12686 (1996).

25. Duyster, J., Bai, R. Y. & Morris, S. W. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20,** 5623–5637 (2001).
26. Onishi, M. *et al.* Applications of retrovirus-mediated expression cloning. *Exp. Hematol.* **24,** 324–329 (1996).
27. Yamashita, Y. *et al.* Sak serine/threonine kinase acts as an effector of Tec tyrosine kinase. *J. Biol. Chem.* **276,** 39012–39020 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Moriuchi for suggestions.

Author Contributions M.S. and Y.L.C. contributed equally to this work. M.S., S.-i.F., H.W. and H.H. constructed the cDNA library and screened for transforming genes. Y.L.C. sequenced the *EML4–ALK* cDNA and conducted the experiments with BA/F3 cells. Y.Y. and S.T. searched for *EGFR* and *KRAS* mutations. M.E., S.I., K.K., M.B., S.O., S.T., Y.I. and H.A. performed RT–PCR for *EML4–ALK* transcripts in cancer specimens. T.N., Y. Sohara, Y. Sugiyama and H.M. designed the overall project, and H.M. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequences of *EML4–ALK* variant 1 and variant 2 cDNA have been deposited in DDBJ, EMBL and GenBank under the accession numbers AB274722 and AB275889, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M. (hmano@jichi.ac.jp).

METHODS

Focus formation assay with a retroviral library. A recombinant retroviral library was constructed as described[7–10] with minor modifications. In brief, total RNA was extracted from a lung cancer specimen isolated from a 62-yr-old man, who gave informed consent. This study was approved by the ethics committee of Jichi Medical University. First-strand cDNA was synthesized from the RNA using PowerScript reverse transcriptase, the SMART IIA oligonucleotide and CDS primer IIA (all from Clontech). The resulting cDNAs were then amplified by PCR with 5′-PCR primer IIA (Clontech) and PrimeSTAR HS DNA polymerase (Takara Bio) for 17 cycles of 98 °C for 10 s and 68 °C for 6 min. The PCR products were ligated to a BstXI adaptor (Invitrogen) and then incorporated into the pMXS retroviral plasmid[26] (provided by T. Kitamura). Recombinant retroviruses were produced by introduction of the plasmid library into the packaging cell line BOSC23 and were used to infect 3T3 cells in the presence of polybrene (4 μg ml^{-1}, Sigma). The cells were cultured for 2 weeks, after which transformed foci were isolated, expanded and subjected to extraction of genomic DNA. Insert cDNAs were recovered from the genomic DNA by PCR with 5′-PCR primer IIA and PrimeSTAR HS DNA polymerase. Amplified products were then ligated to the plasmid pT7Blue-2 (Novagen) and subjected to nucleotide sequencing.

Detection of *EML4–ALK*. To detect a fusion transcript derived from *EML4* and *ALK*, we performed RT with an oligo(dT) primer and total RNA isolated from NSCLC specimens followed by PCR with the primers Fusion-RT-S (5′-GTGCAGTGTTTAGCATTCTTGGGG-3′) and Fusion-RT-AS (5′-TCTTGCCAGCAAAGCAGTAGTTGG-3′) and with a QuantiTect SYBR Green PCR kit (Qiagen). We used the PCR primers 5′-GTCAGTGGTGGACCTGACCT-3′ and 5′-TGAGCTTGACAAAGTGGTCG-3′ for glyceraldehyde-3-phosphate dehydrogenase cDNA, and the primers 5′-TCTGCATTGGAGAGAACAATGTGA-3′ and 5′-TATTCTCCAGTCTCTCTGGGTGGA-3′ for *ALK* cDNA. The *EML4–ALK* fusion gene was detected by PCR, using genomic DNA of the clinical specimen from patient 33 as the template. Amplification was performed with LA Taq DNA polymerase (Takara Bio) and the primers Fusion-genome-S (5′-CCACACCTGGGAAAGGACCTAAAG-3′) and Fusion-genome-AS (5′-AGCTTGCTCAGCTTGTACTCAGGG-3′).

Transforming potential of EML4–ALK. The coding regions of wild-type *ALK* cDNA (provided by S. Morris), wild-type *EML4* cDNA (obtained by RT–PCR) or *EML4–ALK* cDNA were fused to that for the Flag epitope tag (Eastman Kodak) and inserted into the pMXS plasmid. Replacement of the codon for Lys 589 of EML4–ALK with that for Met was performed using a Quickchange Site-Directed Mutagenesis kit (Stratagene). Internal deletion mutants of *EML4–ALK* cDNA were constructed using an ExSite Mutagenesis kit (Stratagene) and the primers 5′-TCGTGACTCAAGAGCTGACAGGCG-3′ and 5′-ATTCGAGCATCACCTTCTCCCCAG-3′ for ΔBasic, 5′-TGACATCTTTATGCTTGTCTGCAG-3′ and 5′-ATTATGAGGAGAGAACTCAGCGAC-3′ for ΔHELP, and 5′-GTGTCGCTGAGTTCTCTCCTCATA-3′ and 5′-GGTGGAGTCATGCTTATATGGAGC-3′ for ΔWD. An expression vector for Flag-tagged NPM–ALK was provided by T. Yamamoto. The various expression plasmids were introduced into 3T3 cells (American Type Culture Collection) by transfection according to the calcium phosphate method, and the cells were then either cultured for 21 d or injected subcutaneously into nude mice. For analysis of dimerization of EML4–ALK, expression vectors for Flag- or Myc-epitope-tagged EML4–ALK or its mutants were introduced into HEK 293 cells, and cell lysates were subsequently subjected to immunoprecipitation with antibodies to Flag (Eastman Kodak) or to Myc (Roche Diagnostics). The resulting precipitates were subjected to immunoblot analysis with these same antibodies. An *in vitro* kinase assay was performed for 30 min as described previously[18] with the synthetic YFF peptide (Operon Biotechnologies Japan). For BA/F3 experiments, cDNAs for ALK, EML4–ALK or EML4–ALK(K589M) were inserted into the plasmid pMX-iresCD8 (ref. 27) to confer simultaneous expression of the protein of interest and mouse CD8. BA/F3 cells were infected with recombinant retroviruses generated from each plasmid, and the resulting CD8-positive cells were purified using a miniMACS cell separation column and magnetic beads conjugated with antibodies to CD8 (both from Miltenyi Biotec). CD8^{+} cells were then incubated with various concentrations of WHI-P154 (EMD Biosciences). Immunoblot analysis of BA/F3 cell lysates was performed with antibodies specific for the Tyr 1604-phosphorylated form of ALK (Cell Signaling Technology).

LETTERS

A ferromagnet in a continuously tunable random field

D. M. Silevitch[1], D. Bitko[2], J. Brooke[3], S. Ghosh[4], G. Aeppli[5] & T. F. Rosenbaum[1]

Most physical and biological systems are disordered, even though the majority of theoretical models treat disorder as a weak perturbation. One particularly simple system is a ferromagnet approaching its Curie temperature, T_C, where all of the spins associated with partially filled atomic shells acquire parallel orientation. With the addition of disorder by way of chemical substitution, the Curie point is suppressed, but no qualitatively new phenomena appear in bulk measurements as long as the disorder is truly random on the atomic scale and not so large as to eliminate ferromagnetism entirely[1]. Here we report the discovery that a simply measured magnetic response is singular above the Curie temperature of a model, disordered magnet, and that the associated singularity grows to an anomalous divergence at T_C. The origin of the singular response is the random internal field induced by an external magnetic field transverse to the favoured direction for magnetization[2–4]. The fact that ferromagnets can be studied easily and with high precision using bulk susceptibility and a large variety of imaging tools will not only advance fundamental studies of the random field problem, but also suggests a mechanism for tuning the strength of domain wall pinning, the key to applications.

Ferromagnets display many interesting effects when exposed to external magnetic fields H. Most important for applications ranging from motors to disk drive read heads is the switching and hysteresis that occurs when the magnetic field changes sign. Magnetic domain-wall pinning makes the magnetization M evolve continuously around an H-dependent loop rather than simply jumping at $H = 0$, as expected without pinning. For fundamental statistical mechanics, equally important are the temperature dependence of M above T_C for an infinitesimal field dH and the H-dependence of M at $T = T_C$. For many ferromagnets, these are both singular, in that they can be described by power laws in the reduced temperature $T - T_C$ and the field H, respectively. Thus, the magnetic susceptibility $\chi = \mathrm{d}M/\mathrm{d}H$ diverges like $(T - T_C)^{-\gamma}$ for small H and like $H^{1/\delta - 1}$ for $T = T_C$. In the last century, these divergences drove vast experimental and theoretical efforts, leading to the formulation of the concept of scale invariance at second order phase transitions[5].

Real-world magnets depart from the ideal system treated by statistical mechanics, typically being riddled with structural and chemical defects that are often deliberately introduced to engineer specific properties. Recognizing this, theorists have produced many models incorporating simplified representations of the disorder seen in realistic magnets. One useful generic model is the 'random-field model', where in addition to a uniform external field, there is a field that varies randomly from site to site with zero mean. Site-random fields are difficult to tune experimentally, and the random field model seemed quite abstract and not amenable to quantitative test until Aharony and Fishman proposed[6] that site-random antiferromagnets in tunable, uniform external magnetic fields should behave like ferromagnets in tunable random fields (Fig. 1a). This insight set off a flurry of activity about 25 years ago[7–12] because insulating antiferromagnets can be characterized optically and by magnetic neutron scattering. Important results were that for small random fields, high dimensional (d) magnets retain their order, and that under these circumstances, the critical exponent γ changes from its clean limit value of 1.25 (ref. 13) to 1.58 for $d = 3$ (ref. 14).

Recently, various authors[2–4] have emphasized that there are also ferromagnets for which an external d.c. field generates an effective internal random field (Fig. 1a). This provides a new venue for investigating the random field problem, with the advantages of a more obvious relation to technical ferromagnetism, the availability of direct bulk probes such as magnetization and susceptibility, and the ability to measure directly the behaviour of $\chi(H)$ for $T \geq T_C$. It is important to note that this direct approach is inapplicable to antiferromagnets because of the inability to apply a staggered (that is, regularly alternating in space) field.

The starting point is the Ising model in a transverse field for a disordered magnet, with hamiltonian $H = -\sum_{\langle ij \rangle} J_{ij} S_i^z S_j^z - \Gamma \sum_i S_i^x$, where S_i is the spin operator at site i and J_{ij} represents the interaction between spins at sites i and j. z is the easy (Ising) axis of magnetization, perpendicular to the magnetic field applied along x. The transverse field Γ leads to quantum mechanical mixing of the Ising up and down spin states at each site, with the outcome that eventually there is a well-known quantum critical point at $T = 0$ for $\Gamma \approx J_{ij}$ beyond which magnetic order along **z** ceases to survive (see Fig. 1b). Γ also generates a non-vanishing expectation value $\langle S_i^x \rangle$ at each site, which will exert a field on the Ising components S_j^z of the spin operator if an interaction with off-diagonal (in spin space) terms is added to the hamiltonian and these terms do not cancel by symmetry. An example of such an interaction is the magnetic dipolar coupling for a ferromagnet simply disordered by site dilution, illustrated in Fig. 1a.

For our experiments on the new route to the random field Ising model, we use $LiHo_xY_{1-x}F_4$, a transparent insulator. Reference 15 contains a recent description of the underlying hamiltonian, including crystal fields, a laboratory magnetic field H, and nuclear hyperfine interactions, and their detailed consequences for the magnetic dynamics. The Ho^{3+} ions carry large magnetic moments, strongly aligned along the **c** axis of the tetragonal material and coupled to each other via the dipolar interaction. The outcome is that pure $LiHoF_4$ is an effective spin-1/2 Ising ferromagnet with a Curie temperature $T_C = 1.53$ K (ref. 16). Fields H_t transverse to **c** convert the pure compound into a realization of the transverse field Ising model, with a continuous phase boundary between the $H_t = 0$ classical critical point and a $T = 0$ quantum critical point at $H_t = 50$ kOe (ref. 17). For $H_t < 20$ kOe, Γ is quadratic in H_t (ref. 18). Mean field theory, taking into account both the electronic and nuclear spin degrees of freedom, quantitatively describes the static and dynamical properties in the H_t–T plane, including the classical ($H_t \rightarrow 0$) and quantum ($T \rightarrow 0$) limits[15,17,19]. When non-magnetic yttrium partially replaces magnetic holmium, the classical disordered ferromagnet orders at a

[1]The James Franck Institute and Department of Physics, The University of Chicago, Chicago, Illinois 60637, USA. [2]MPI Research Inc., East Brunswick, New Jersey 08816, USA. [3]Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts 02420, USA. [4]School of Natural Sciences, University of California Merced, Merced, California 95344, USA. [5]London Centre for Nanotechnology and Department of Physics and Astronomy, UCL, London WC1E 6BT, UK.

depressed $T_C(x) = xT_C(x=1)$ down to $x \approx 0.2$, where the combination of disorder and frustration arising from the spatial anisotropy of the dipole interaction gives rise to a spin glass ground state[1].

Whereas the measured Curie point has precisely the value anticipated in the simplest molecular field theory, the quantum critical point occurs at greatly suppressed values, as can be seen in the phase diagrams (two of which have been published previously[17,20]) of Fig. 1b. Indeed, it was the discrepancy between simple theory and these data that motivated the two recent papers[2,3] that pointed out the generation of internal random fields in disordered dipole-coupled Ising magnets. In this work, we focus on the region near and above the ferromagnetic $T_C(x)$ to explore the random-field induced change in the classical, 'high-temperature' critical behaviour. By contrast, our previous work on the low-temperature quantum behaviour of the same materials focused on different phenomena, including the quantum glass[18], entanglement (caused by the same off-diagonal terms in the dipole interaction as are responsible for the random fields discussed here)[4] and antiglass behaviour[21], decoherence[15], tunable domain wall tunnelling[22] and the associated concept of quantum annealing[20].

Figure 2 shows the evolution of the real part of the magnetic susceptibility, $\chi'(\Gamma)$, (see Methods) with decreasing T for the most disordered of our ferromagnetic samples, namely the $x = 0.44$ crystal. At high T, the susceptibility is sharply peaked, and its divergence at the Curie point is only cut off by demagnetizing effects due to sample geometry. The extraordinary sharpness of the susceptibility peaks again confirms the ideal nature of $LiHo_xY_{1-x}F_4$ as a model system for studying the combined effects of quantum mechanics and quenched disorder. For $T < 0.300$ K, however, the peak becomes significantly broader and rounder, and the peak susceptibility never reaches the demagnetization limit. By contrast, comparable measurements on pure $LiHoF_4$ reach the demagnetization limit for temperatures as low as 0.025 K (refs 17, 23). As we have shown previously[20,22], the system becomes a randomly pinned ferromagnetic domain state and in many respects behaves like a glass, unable to reach equilibrium over the measurement time, yet retains a net moment as indicated by local Hall-probe magnetometry. These results are in excellent qualitative agreement with the concept that with growing transverse field, the induced random field becomes large enough to prevent the achievement of an equilibrium ferromagnetic state, in exact analogy with what had previously been found for the random Ising antiferromagnets in external fields.

We now focus on the critical behaviour for $x = 1$ and $x = 0.44$ on the approach to the classical critical point $T_C(\Gamma{=}0)$. Figure 3 shows the susceptibility for the two concentrations near $T_C(\Gamma{=}0)$ as a function of T at $\Gamma = 0$, as well as Γ at $T = T_C$. Both the ordered ($x = 1$) and disordered ($x = 0.44$) crystals show power-law behaviour, $\chi' \propto T^{-\gamma}$, over multiple decades as a function of temperature. In both cases, the thermal critical exponent was measured to be $\gamma = 1.00 \pm 0.04$, in accordance with the mean-field prediction[24–26]. As a function of Γ, the ordered system shows an apparent critical exponent of $1/\delta - 1 = -0.99 \pm 0.0005$. The low-$\Gamma$ asymptote is to be contrasted with $\chi' \propto \Gamma^{-2}$, given by mean field theory for the transverse field Ising model when the hyperfine interactions are either negligible or much larger than Γ. For Ho^{3+}, the hyperfine coupling $A = 39$ mK (ref. 15), and because the nuclear and electronic spins are large (7/2 and 8 respectively), the bandwidth of nuclear excitations is therefore also large. This accounts for the bending of the susceptibility towards very shallow behaviour as Γ goes beyond A. At the same time, for the smallest Γ, we will be dealing with both absolute uncertainties and fluctuations in the temperature, meaning that the expected $1/\Gamma^2$ singularity will be smoothed out to yield the less singular, but still strongly divergent $1/\Gamma$ form that we actually observe.

The disordered system demonstrates qualitatively different behaviour (Fig. 3b). Nearest the classical critical point, we observe power-law behaviour with an exponent of $1/\delta - 1 = -0.57 \pm 0.03$, with a crossover at larger Γ to $1/\delta - 1 = -1$. We emphasize that these are entirely intrinsic results, and although the hyperfine interactions clearly remain a factor, thermal broadening is not an issue here as it is for very small Γ for the pure compound, where the phase boundary rises very steeply. There is a more remarkable result at temperatures above T_C. Figure 4 reveals non-analytic behaviour of $\chi'(H_t)$, in that there is no rounding as $H_t \to 0$ (note that χ', measured along **c**, is an even function of H_t, peaked at $H_t = 0$); a prior study on a more dilute and non-ferromagnetic ($x = 0.17$) concentration observed a hint of a cusp at $H_t = 0$ rather than a rounded maximum at $H_t = 0$

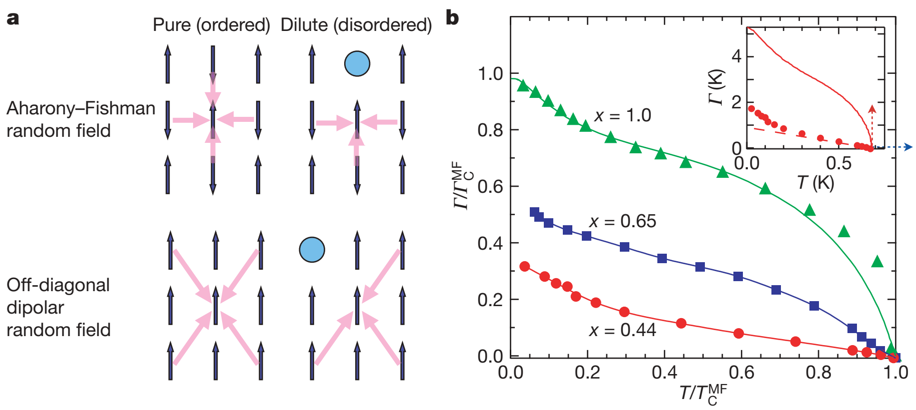


Figure 1 | Random fields in a diluted, dipolar-coupled ferromagnet drive the system away from mean-field behaviour. a, Diagram showing two realizations of the random-field model. Top row, Aharony-Fishman method[2], using a site-diluted antiferromagnet. Bottom row, dilute dipole-coupled Ising ferromagnet[11–13], where the off-diagonal terms of the dipole interaction act to enable the random fields. In both cases, the undiluted system (left column) experiences no random-field effects due to overall symmetry; breaking the symmetry by dilution with non-magnetic sites (right column) introduces a net random field. **b**, Normalized ferromagnetic–paramagnetic phase diagram for $LiHo_xY_{1-x}F_4$. From top to bottom, x = 1.0, 0.65 and 0.44, with $T_C = 1.53$, 1.02 and 0.669 K, respectively. For $x = 1$, the solid line is derived from mean field theory[15] including nuclear and electronic degrees of freedom, whereas for the other compositions, the curves are guides for the eye. The normalization constants T_C^{MF} and Γ_C^{MF} are the values $xT_C(x{=}1)$ and $\Gamma_C = x\Gamma_C(x{=}1)$ predicted from mean-field theory at $\Gamma = 0$ and $T = 0$, respectively. The upturn for $T < 0.4$ K reflects the influence of the hyperfine interaction where the coupling between electronic (J) and nuclear (I) spins creates a larger effective moment ($I + J$) Inset, phase boundary for $x = 0.44$ in absolute units. Solid line shows the phase boundary predicted by mean field theory, using the parameters derived from the model for $x = 1$, with the average spin–spin coupling strength J scaled by the concentration. Dashed line shows the phase boundary derived from the critical divergences (see text for details). Dashed lines with arrows are where the critical curves plotted in Fig. 3b were measured.

(ref. 17). Mathematically, this means that we must add an anomalous linear term $a|H_t|$ to the expansion $\chi(H_t) = \chi(T,H_t = 0) + bH_t^2 + \ldots$. That the cusp at $H_t = 0$ is a maximum rather than a minimum implies that the coefficient a of the new term must be negative, as is that of the quadratic term. There is consequently an added suppression of the Ising correlations due to the disorder, and it is precisely given by a term of first order in the amplitude $|H_t|$ of the random field generated by H_t.

The $T > T_C$ singularity $\chi' \approx \chi'_0 - a|H_t|$, has the same leading order (in $|H_t|$) behaviour as the Curie–Weiss-like expression $\chi' \approx 1/(c_1 + c_2|H_t|)$, which itself is consistent with the measured exponent $\gamma \approx 1/2$ found at T_C (when $c_1 = 0$) because $\Gamma \propto H_t^2$. We therefore can encapsulate both the $T > T_C$ singularity and the anomalous exponent at T_C in the following modified Curie law for the susceptibility per Ho atom:

$$\chi' = \frac{C}{\alpha' \mu_B |H_t| + (T - T_C) + \gamma' \Gamma} \tag{1}$$

where the term proportional to Γ takes account of the quantum fluctuations in the disordered system as well as higher-order random field effects, and C is the Curie constant for an isolated Ho^{3+} ion in the fluoride. The parameters $\alpha'_{0.44} = 0.157 \pm 0.001$ and $\gamma'_{0.44} = 0.321 \pm 0.001$ were determined by fitting $\chi'(\Gamma)$ at $T = T_C$ to equation (1), as shown in Fig. 3b.

As shown in the inset of Fig. 4, the measured susceptibilities for temperatures above and below T_C collapse onto a single universal curve following equation (1), using the values of α' and γ' determined from Fig. 3b. Indeed, to within experimental accuracy, equation (1) is an exact description of the susceptibility at the classical critical point, with $T_C = 0.669$ K and $H_t = 0$ for $x = 0.44$. The loci in Γ–T space where equation (1) diverges are also consistent with the direct measurement of the phase boundary shown in Fig. 1, which is linear in Γ from 0.2 K to 0.625 K, with a crossover to linear in $|H_t|$ approaching the classical critical point.

Our experiments are to our knowledge the first to approach the critical behaviour of the random field Ising problem in a ferromagnet where the control parameter is the random field amplitude itself. Over many decades of reduced temperature and random field amplitude, a generalized Curie–Weiss form, equation (1), describes the data. This is perhaps not surprising, given that Curie–Weiss descriptions inevitably arise in molecular field theories applicable when interactions are long ranged, like the dominant dipolar interaction in $Li(Ho,Y)F_4$. What is unexpected, however, is that the leading order contribution of the random field term is $|H_t|$ rather than H_t^2; any perturbative approaches that we have used to determine the random field effects on χ yield only even order terms in H_t. The only means for generating the linear term are for rare, large amplitude random fields to become so important that they destroy the viability of perturbation expansions in the product of the off-diagonal interaction term and H_t. Our detailed results for χ, therefore, are a dramatic manifestation of 'Griffiths singularities'[27]. A particularly simple outcome of the Griffiths singularities is the shape of the phase diagram $T_C(\Gamma)$ terminating at the quantum critical point ($T = 0$ and $\Gamma = \Gamma_C$), which does not agree with mean field theory for $\Gamma > 0$. The divergence of χ' in equation (1) for our $x = 0.44$ samples occurs at a line of Curie points where $T_C(H_t) - T_C(H_t = 0) = -\alpha' \mu_B |H_t| - \gamma' \Gamma = -0.16\mu_B |H_t| - 0.32\Gamma$, implying a low field asymptote of $T_C(H_t) - T_C(H_t = 0) \propto |H_t|$. For the less dilute sample ($x = 0.65$), it appears that $T_C(H_t) - T_C(H_t = 0) \propto H_t^2$, in agreement with perturbation theory. The numerical results of ref. 2, while covering a much coarser grid in the H_t–T plane than our experiments, are consistent with the appearance of the linear term with larger dilution.

We have implemented a method to execute quantitative studies of the long-established random field problem, and have made surprising discoveries using conventional magnetic susceptometry. This opens the door to investigation of the random field problem by other techniques, many of which, such as magnetic noise spectroscopy, magneto-optics and magnetic force microscopy, are much better matched to ferromagnets than antiferromagnets. This will result in the ability to examine both the statics and dynamics of the random-field problem, and perform detailed checks of long-standing theoretical predictions.

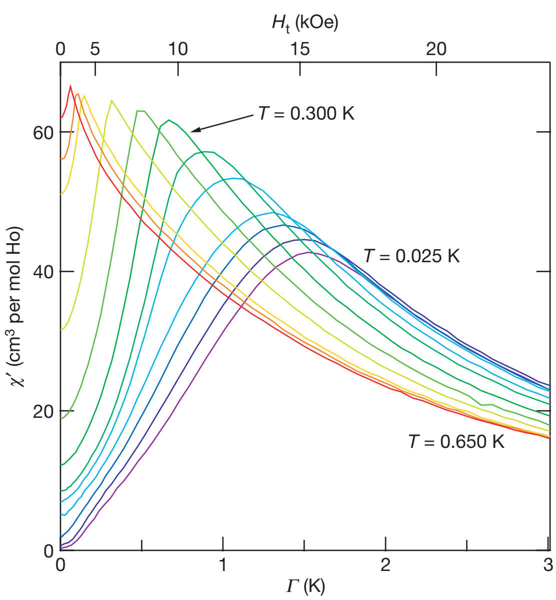


Figure 2 | Real part of the magnetic susceptibility χ' of $LiHo_{0.44}Y_{0.56}F_4$ measured versus transverse field Γ at a series of temperatures. (Here T = 0.650, 0.625, 0.600, 0.500, 0.400, 0.300, 0.200, 0.150, 0.100, 0.080, 0.060 and 0.025 K.) Below T = 0.300 K, the cusp in χ' becomes rounded at the Curie point and the response no longer reaches the demagnetization limit as glassy effects set in.

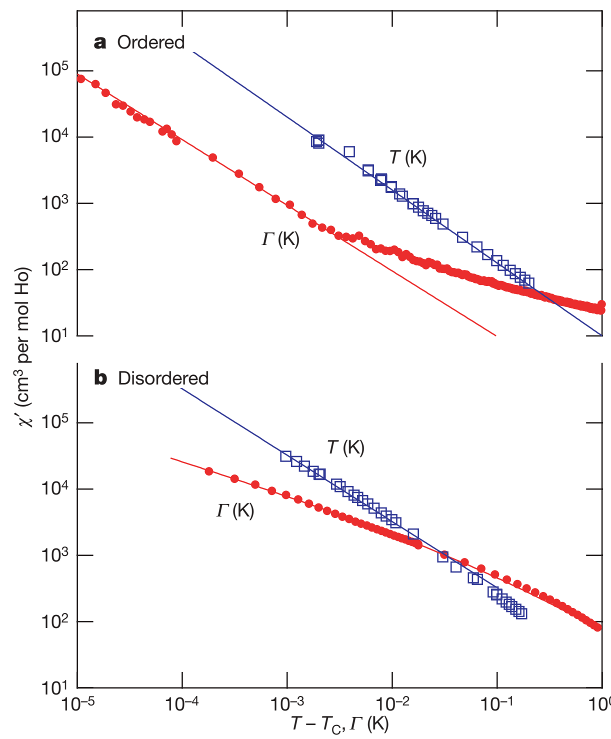


Figure 3 | Critical behaviour of $LiHo_xY_{1-x}F_4$ in the paramagnetic regime as a function of T at $\Gamma = 0$ and Γ at $T = T_C$. Lines are power law fits described in the text. **a**, $x = 1.00$ ($T_C = 1.53$ K); **b**, $x = 0.44$ ($T_C = 0.669$ K). $\chi(T)$ shows mean-field behaviour, whereas $\chi(\Gamma)$ exhibits a suppressed critical exponent indicative of the effects of the random fields, fitted to equation (1) in the text. For both concentrations, absolute error in T_C is ± 0.2 mK with a thermal stability of ± 0.02 mK.

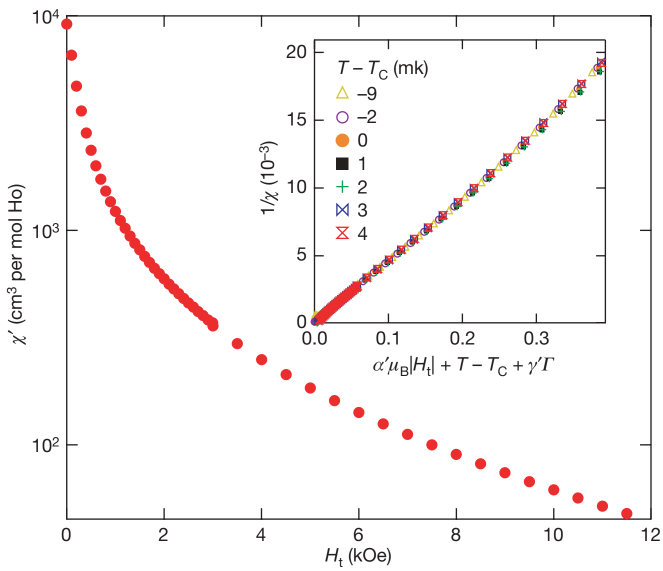


Figure 4 | Singularities above the Curie temperature. Main plot, singular behaviour of the susceptibility of $LiHo_{0.44}Y_{0.56}F_4$ at $T = 0.673$ K (0.004 K above T_C). Inset, inverse susceptibility at a series of temperatures above, at and below T_C, collapsed onto a universal, singular curve using the fitting parameters derived in Fig. 3. See text for details.

METHODS

We performed a.c. susceptibility measurements on single crystals of $LiHo_xY_{1-x}F_4$ with $x = 1.0$, 0.65 and 0.44. The $(5 \times 5 \times 10)$ mm^3 single-crystal samples with long axis along **c** were mounted on the cold finger of a helium dilution refrigerator placed inside a 8 T superconducting magnet with field transverse to the Ising axis. AC excitation fields of 20–50 mOe at 10 Hz (chosen to ensure linear response) were applied along the Ising axis with the magnetic response recorded using an inductive pickup coil in a gradiometer configuration.

Received 12 March; accepted 19 June 2007.

1. Reich, D. H. *et al.* Dipolar magnets and glasses: Neutron-scattering, dynamical, and calorimetric studies of randomly distributed Ising spins. *Phys. Rev. B* **42,** 4631–4644 (1990).
2. Tabei, S. M. A., Gingras, M. J. P., Kao, Y.-J., Stasiak, P. & Fortin, J.-Y. Induced random fields in the $LiHo_xY_{1-x}F_4$ quantum Ising magnet in a transverse magnetic field. *Phys. Rev. Lett.* **97,** 237203 (2006).
3. Schechter, M. $LiHo_xY_{1-x}F_4$ as a random field Ising ferromagnet. Preprint at ⟨http://arxiv.org/cond-mat/0611063⟩ (2006).
4. Ghosh, S., Rosenbaum, T. F., Aeppli, G. & Coppersmith, S. N. Entangled quantum state of magnetic dipoles. *Nature* **425,** 48–51 (2003).
5. Ma, S. K. *Modern Theory of Critical Phenomena* (Addison-Wesley, Reading, Massachusetts, 1976).
6. Fishman, S. & Aharony, A. Random field effects in disordered anisotropic anti-ferromagnets. *J. Phys. C* **12,** L729–L733 (1979).
7. Gofman, M. *et al.* Critical behavior of the random-field Ising model. *Phys. Rev. B* **53,** 6362–6384 (1996).
8. Yoshizawa, H. *et al.* Random-field effects in two- and three-dimensional Ising antiferromagnets. *Phys. Rev. Lett.* **48,** 438–441 (1982).
9. Ferreira, I. B., King, A. R., Jaccarino, V., Cardy, J. L. & Guggenheim, H. J. Random-field-induced destruction of the phase transition of a diluted two-dimensional Ising antiferromagnet: $Rb_2Co_{0.85}Mg_{0.15}F_4$. *Phys. Rev. B* **28,** 5192–5198 (1983).
10. Belanger, D. P., King, A. R., Jaccarino, V. & Cardy, J. L. Random-field critical behavior of a d=3 Ising system. *Phys. Rev. B* **28,** 2522–2526 (1983).
11. Birgeneau, R. J. *et al.* Critical behavior of a site-diluted three-dimensional Ising magnet. *Phys. Rev. B* **27,** 6747–6753 (1983).
12. Belanger, D. P. & Young, A. P. The random field Ising model. *J. Magn. Magn. Mater.* **100,** 272–291 (1991).
13. Belanger, D. P. & Yoshizawa, H. Neutron scattering and the critical behavior of the three-dimensional Ising magnet FeF_2. *Phys. Rev. B* **35,** 4823–4830 (1987).
14. Slanic, Z., Belanger, D. P. & Fernandez-Baca, J. A. Equilibrium random-field Ising critical scattering in the antiferromagnet $Fe_{0.93}Zn_{0.07}F_2$. *Phys. Rev. Lett.* **82,** 426–429 (1999).
15. Ronnow, H. M. *et al.* Quantum phase transition of a magnet in a spin bath. *Science* **308,** 389–392 (2005).
16. Hansen, P. E., Johansson, T. & Nevald, R. Magnetic properties of lithium rare-earth fluorides: Ferromagnetism in $LiErF_4$ and $LiHoF_4$ and crystal-field parameters at the rare-earth and Li sites. *Phys. Rev. B* **12,** 5315–5324 (1975).
17. Bitko, D., Rosenbaum, T. F. & Aeppli, G. Quantum critical behavior for a model magnet. *Phys. Rev. Lett.* **77,** 940–943 (1996).
18. Wu, W., Ellman, B., Rosenbaum, T. F., Aeppli, G. & Reich, D. H. From classical to quantum glass. *Phys. Rev. Lett.* **67,** 2076–2079 (1991).
19. Chakraborty, P. B., Henelius, P., Kjonsberg, H. & Sandvik, A. W. &. Girvin, S. M. Theory of the magnetic phase diagram of $LiHoF_4$. *Phys. Rev. B* **70,** 144411 (2004).
20. Brooke, J., Bitko, D., Rosenbaum, T. F. & Aeppli, G. Quantum annealing of a disordered magnet. *Science* **284,** 779–781 (1999).
21. Ghosh, S., Parthasarathy, R., Rosenbaum, T. F. & Aeppli, G. Coherent spin oscillations in a disordered magnet. *Science* **296,** 2195–2198 (2002).
22. Brooke, J., Rosenbaum, T. F. & Aeppli, G. Tunable quantum tunnelling of magnetic domain walls. *Nature* **413,** 610–613 (2001).
23. Bitko, D. *Order and Disorder in a Model Quantum Magnet.* Ph.D. thesis, Univ. Chicago (1997).
24. Aharony, A. & Halperin, B. I. Exact relations among amplitudes at critical points of marginal dimensionality. *Phys. Rev. Lett.* **35,** 1308–1310 (1975).
25. Brezin, E. & Zinn-Justin, J. Critical behavior of uniaxial systems with strong dipolar interactions. *Phys. Rev. B* **13,** 251–254 (1976).
26. Ahlers, G., Kornblit, A. & Guggenheim, H. J. Logarithmic corrections to the Landau specific heat near the Curie temperature of the dipolar Ising ferromagnet $LiTbF_4$. *Phys. Rev. Lett.* **34,** 1227–1230 (1975).
27. Griffiths, R. B. Nonanalytic behavior above the critical point in a random Ising ferromagnet. *Phys. Rev. Lett.* **23,** 17–19 (1969).

Acknowledgements The work at the University of Chicago was supported by the US Department of Energy and the MRSEC programme of the US National Science Foundation, while work in London was supported by the UK Engineering and Physical Sciences Research Council and a Wolfson-Royal Society Research Merit Award.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.F.R. (tfr@uchicago.edu).

Vol 448 | 2 August 2007 | doi:10.1038/nature06037

LETTERS

Electronic spin transport and spin precession in single graphene layers at room temperature

Nikolaos Tombros[1], Csaba Jozsa[1], Mihaita Popinciuc[2], Harry T. Jonkman[2] & Bart J. van Wees[1]

Electronic transport in single or a few layers of graphene is the subject of intense interest at present. The specific band structure of graphene, with its unique valley structure and Dirac neutrality point separating hole states from electron states, has led to the observation of new electronic transport phenomena such as anomalously quantized Hall effects, absence of weak localization and the existence of a minimum conductivity[1]. In addition to dissipative transport, supercurrent transport has also been observed[2]. Graphene might also be a promising material for spintronics and related applications, such as the realization of spin qubits, owing to the low intrinsic spin orbit interaction, as well as the low hyperfine interaction of the electron spins with the carbon nuclei[3,4]. Here we report the observation of spin transport, as well as Larmor spin precession, over micrometre-scale distances in single graphene layers. The 'non-local' spin valve geometry was used in these experiments, employing four-terminal contact geometries with ferromagnetic cobalt electrodes making contact with the graphene sheet through a thin oxide layer. We observe clear bipolar (changing from positive to negative sign) spin signals that reflect the magnetization direction of all four electrodes, indicating that spin coherence extends underneath all of the contacts. No significant changes in the spin signals occur between 4.2 K, 77 K and room temperature. We extract a spin relaxation length between 1.5 and 2 μm at room temperature, only weakly dependent on charge density. The spin polarization of the ferromagnetic contacts is calculated from the measurements to be around ten per cent.

The elementary device geometry to detect spin transport is a two-terminal spin valve in which a non-magnetic medium is contacted by two ferromagnetic electrodes[5]. Hill *et al.*[6] have studied devices with 200 nm-spaced transparent permalloy contacts to a single graphene layer. Relatively large magnetoresistances of several hundred ohms were observed in the magnetic-field region in which the magnetization direction switched but no clear distinction between parallel and anti-parallel configurations could be observed. These experiments suggest the possibility of spin-dependent transport, but the potential contribution of spurious effects has yet to be excluded.

In our experiments we adopt the four-terminal 'non-local' technique[7,8]. Here the charge current path can be fully separated from the voltage detection circuit. As explained in Fig. 1, if spin injection and spin transport takes place, a bipolar spin signal should be observed that changes sign when the magnetization configuration of the electrodes switches from parallel to anti-parallel. The absence of a background resistance means that the non-local technique is less sensitive to device resistance fluctuations and spurious magnetoresistances (such as Hall effects), compared to the standard two-terminal spin valve technique. The non-local technique has been applied to metals[7,8], semiconductors[9] and carbon nanotubes[10].

As described in the Methods, we failed to observe spin transport in graphene devices with transparent low ohmic ferromagnetic contacts. Our experiments require the use of tunnel barriers between the ferromagnetic electrodes and the graphene layer underneath, to increase the spin-dependent interface resistance and to combat the conductivity mismatch problem[11–13]. Here the presence of a thin Al_2O_3 layer should (at least in principle) create a spin-dependent tunnel barrier. Also, it is essential for our devices, which have electrodes that completely overlap the graphene strips, that the barrier should be transparent enough to result in measurable ($<1\,\mathrm{M\Omega}$) resistances, but at the same time opaque enough that carriers can pass underneath it with conservation of spin direction (Fig. 1c, d). Our measurements show that we have achieved this. However, given the expected thickness of the oxide of 0.8 nm it is unlikely that the tunnel barrier will be uniform, and pinholes may be expected.

The fabrication of the devices is described in the Methods. Figure 1a shows the device geometry and Fig. 1b the measurement set-up. We prepared cobalt electrodes with widths from 80 to 800 nm. Their coercive fields depend on their width and range from 150 mT down to 30 mT. Although there is some scatter in the coercive fields for similar contact widths, we can always reconstruct which contact is switching at what magnetic field from the measurements of the various spin valve signals[14]. The non-local resistance is recorded while the magnetic field (applied in the y direction) is swept from a negative value to a positive value, followed by an opposite sweep back to negative values. The electrical measurements are performed with standard low-frequency lock-in techniques. Currents are in the range 100 nA to 5 μA.

Figure 2a shows a typical non-local measurement of device 1 at 4.2 K. Because there are four ferromagnetic electrodes involved, several resistance values are observed that we can associate with a particular configuration of magnetization directions. We note that spin transport over at least 330 nm (the spacing between the centre electrodes) can occur. However, the observation that there are at least three resistance levels indicates that the outer electrodes are also spin injectors and detectors, implying an even longer length scale of 1 μm. This device was the only one in which we could also observe a spin valve signal in the two-terminal 'local' geometry (Fig. 2b). Here the resistance is dominated by the contact resistances that show fluctuations as a function of time, resulting in a noise background. Nevertheless, at the switching fields of the central electrodes 2 and 3 an increase of typically 60 Ω can be observed. This is in agreement with the fact that, if both can be measured, the local signal should be larger than the non-local signal[15]. We note that the identification of these switches as spin signals would have been very difficult without the support of the non-local measurements. Figure 2c shows a comparison between the 4.2 and 77 K data of device 1. Here the magnetization direction of only one of the central electrodes is reversed. This results in a typical 'minor loop' shape, known from measurements on spin valves with electrodes with different coercivities. We observe only a weak dependence on temperature.

[1]Physics of Nanodevices, [2]Molecular Electronics, Zernike Institute for Advanced Materials, Nijenborgh 4, 9747 AG Groningen, The Netherlands.

To investigate the dependence of the signal on the electrode separation at 77 K, we prepared and measured devices 2 and 3, in which graphene strips 2 µm wide were used. On each device a sequence of ferromagnetic contacts with monotonically increasing spacings and different widths were deposited. We will give details elsewhere (manuscript in preparation). We summarize the spin signals as follows: for spacing 350 nm, the resistance was 31 Ω; for 550 nm, 35 Ω; for 1 µm, 21 Ω; for 1.5 µm, 21 Ω; for 2 µm, 22 Ω; for 3 µm, 8 Ω; and for 5 µm, 5 Ω. This shows that the spin signals start to go down at spacings beyond 2 µm, indicating a spin relaxation length of about 2 µm.

Spin transport at room temperature is shown in Fig. 3a for device 4. The magnitude of the spin signal (~6 Ω) is comparable to that observed at 77 K for a 3-µm-spacing device. Figure 3b shows the gate voltage dependence of the non-local signal in parallel and antiparallel configurations. In agreement with the principle of the non-local technique, it shows that the parallel configuration always yields a positive value and the antiparallel a negative one. In Fig. 3c the gate voltage V_g dependence of the graphene resistance is plotted. The typical shape for single graphene layers is observed, with a slightly shifted neutrality point at $V_g = 19$ V, where the resistivity reaches a maximum of 3.2 kΩ (ref. 1). A comparison between Fig. 3c and Fig. 3b shows that there is only a small decrease of the spin signal at the Dirac point.

Next, we present room-temperature Hanle-type spin precession experiments. For this purpose, we first applied a magnetic field in the y direction to prepare the electrodes in a parallel or antiparallel magnetization direction. Then this field was removed and a magnetic field B in the z direction was scanned. The data are shown in Fig. 4. Electrons are injected with a spin polarization in the up direction by contact 3. They precess around the B field with a Larmor frequency $\omega_L = g\mu_B B/\hbar$, with g the effective Landé factor (~2) and μ_B the Bohr

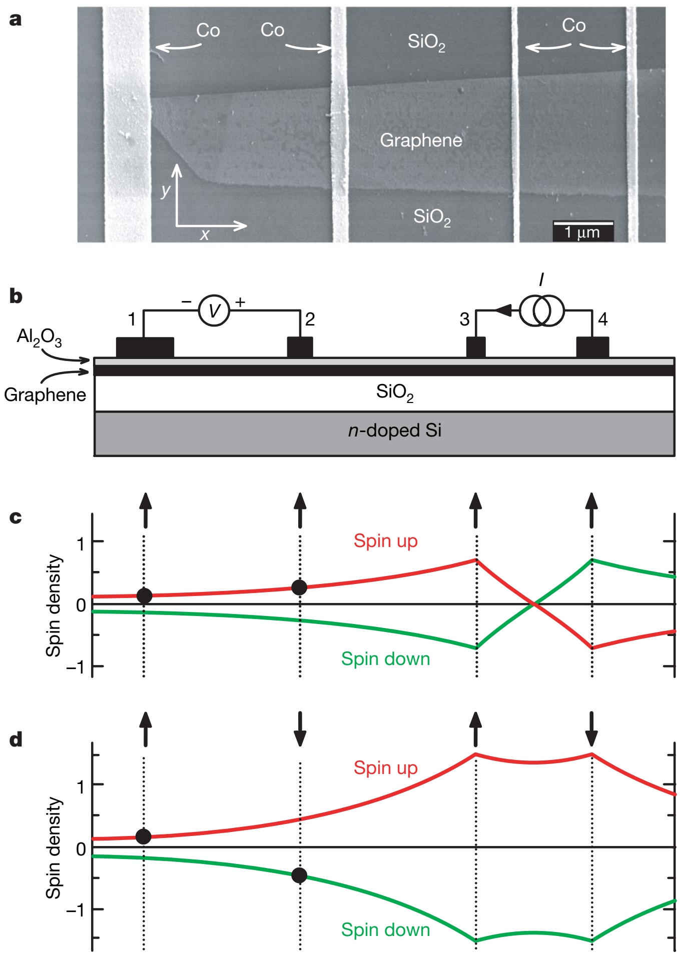


Figure 1 | Spin transport in a four-terminal spin valve device. **a**, Scanning electron micrograph of a four-terminal single-layer graphene spin valve. Cobalt electrodes (Co) are evaporated across a single-layer graphene strip prepared on a SiO_2 surface. **b**, The non-local spin valve geometry. A current I is injected from electrode 3 through the Al_2O_3 barrier into graphene and is extracted at contact 4. The voltage difference is measured between contacts 2 and 1. The non-local resistance is $R_{\text{non-local}} = (V_+ - V_-)/I$. **c**, Illustration of spin injection and spin diffusion for electrodes having parallel magnetizations. Injection of up spins by contact 3 results in an accumulation of spin-up electrons underneath contact 3, with a corresponding deficit of spin-down electrons. Owing to spin relaxation the spin density decays on a scale given by the spin relaxation length. The dots show the electric voltage measured by contacts 1 and 2 in the ideal case of 100% spin selectivity. A positive non-local resistance is measured. (We note that a larger positive signal can be obtained by reversing the magnetization direction of contact 1). **d**, Spin injection and spin diffusion for antiparallel magnetizations. The voltage contacts probe opposite spin directions, resulting in a negative non-local resistance.

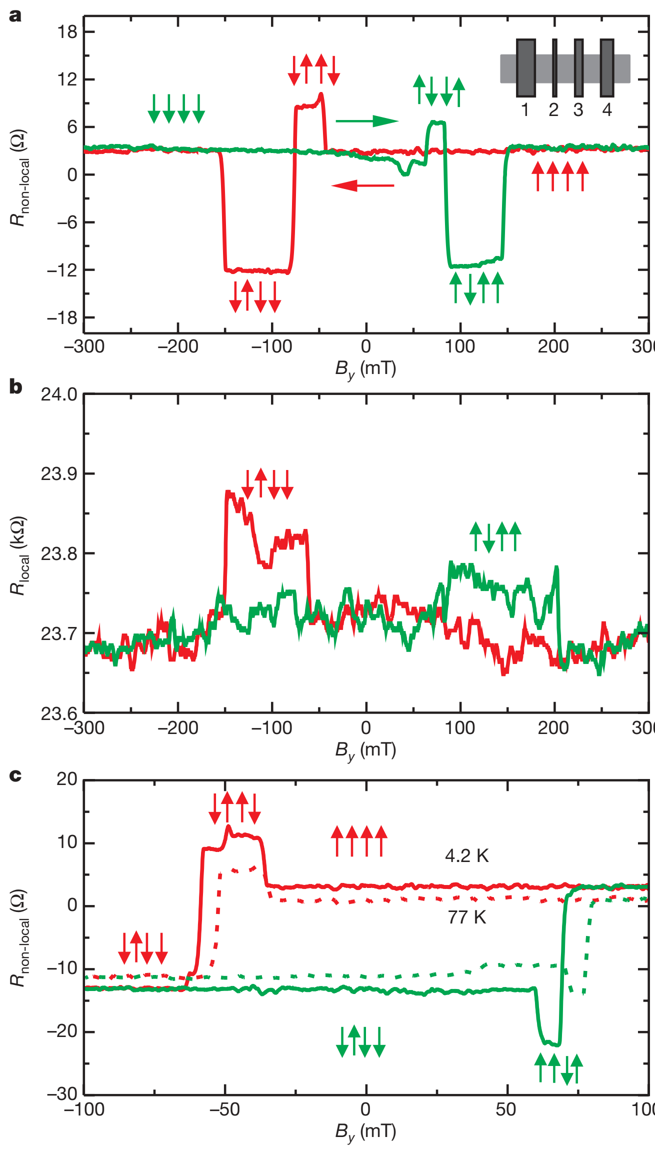


Figure 2 | Spin transport at 4.2 and 77 K. **a**, Non-local spin valve signal for device 1 at 4.2 K. The sweep directions of the magnetic field are indicated (red or green arrows). The magnetic configurations of the electrodes are illustrated for both sweep directions. The widths of the electrodes are 330 nm (for electrode 1), 90 nm (for 2), 140 nm (for 3), 250 nm (for 4), and the electrode spacings were 330 nm for the gaps between electrodes 1 and 2, 2 and 3, and 3 and 4. The graphene width was 1.4 µm. Inset, electrode spacing configuration. **b**, A two-terminal local spin valve signal (measured between contacts 2 and 3) of about 60 Ω is measured at 4.2 K. **c**, Spin signals at 4.2 and 77 K. A 'minor loop' is observed because the magnetic field sweep is reversed before contact 2 has reversed its magnetization direction.

magneton. The data shows that spin-up electrons injected by contact 3 precess while they are diffusing towards contact 2. At $B \approx 100$ mT their average precession angle when they arrive at this contact is 180°, resulting in a sign reversal of the spin signal.

The spin density is constant along the y direction, so we fitted the detailed shape of the data with the one-dimensional Bloch equations, which describe the combined effect of diffusion, precession and spin relaxation[7–9]. From the fitting procedure both the diffusion constant D and the transverse spin relaxation time T_2 can be obtained. As we will see later, the longitudinal relaxation time T_1, extracted from the exponential length dependence of the spin valve signal in sample 5, is similar to T_2 and therefore the use of a single spin relaxation time τ_{sf} is justified. The fits indicate a spin relaxation length, $\lambda_{sf} = \sqrt{D\tau_{sf}}$, between 1.5 and 2 μm, with only minor differences between the high density regime and the Dirac point.

We now calculate D from the measured conductivity σ. At an energy ε sufficiently far away from the Dirac point the density of states for a single graphene layer is given by $\nu(\varepsilon) = g_v g_s 2\pi|\varepsilon|/(h^2 v_F^2)$, which includes the twofold valley ($g_v = 2$) and spin ($g_s = 2$) degeneracy ($v_F \approx 10^6$ m s^{-1}) (ref. 16). Integration yields the total density $n(\varepsilon) = g_v g_s \pi \varepsilon^2/(h^2 v_F^2)$. From the measurements of Fig. 3c, together with the device geometry ($L = 3$ μm and $W = 600$ nm), we obtain the conductivity $\sigma = 1.2 \times 10^{-3}\,\Omega^{-1}$ at $V_g = -40$ V, where $n = 3.6 \times 10^{16}$ m^{-2}.

The Einstein relation is $\sigma = \nu e^2 D$, where the diffusion constant in two-dimensions is given by $D = 0.5 v_F l$, with l the scattering mean free path. We find $D = 1.8 \times 10^{-2}$ m^2 s^{-1} and $l = 36$ nm. This value is very close to the value of D obtained by independent fitting of the data in Fig. 4. This supports the interpretation of the signal as being due to the the Larmor precession of diffusing electron spins.

We use a similar procedure at the Dirac neutrality point ($V_g = 19$ V). Here we start with this expression[4,17]: $\nu(\varepsilon = 0) = 4\pi/(h v_F l)$. Applying the Einstein relation with $\sigma(V_g = 19\,\text{V}) = 0.33 \times 10^{-3}\,\Omega^{-1}$ we find $D = 2.2 \times 10^{-2}$ m^2 s^{-1}. Given the numerical uncertainties, this shows that the diffusion constant is not changed significantly compared to the value at high metallic densities. Although the agreement is less good than at higher densities the comparison with the fitted value in Fig. 4 confirms that the description of simultaneous spin diffusion and precession also applies at the Dirac neutrality point.

The obtained values for the spin signals can now be compared with theory. We modify this three-dimensional description[8] for use in two-dimensions:

$$R_{\text{non-local}} = \frac{P^2 \lambda_{sf}}{2W\sigma} \exp(-L/\lambda_{sf})$$

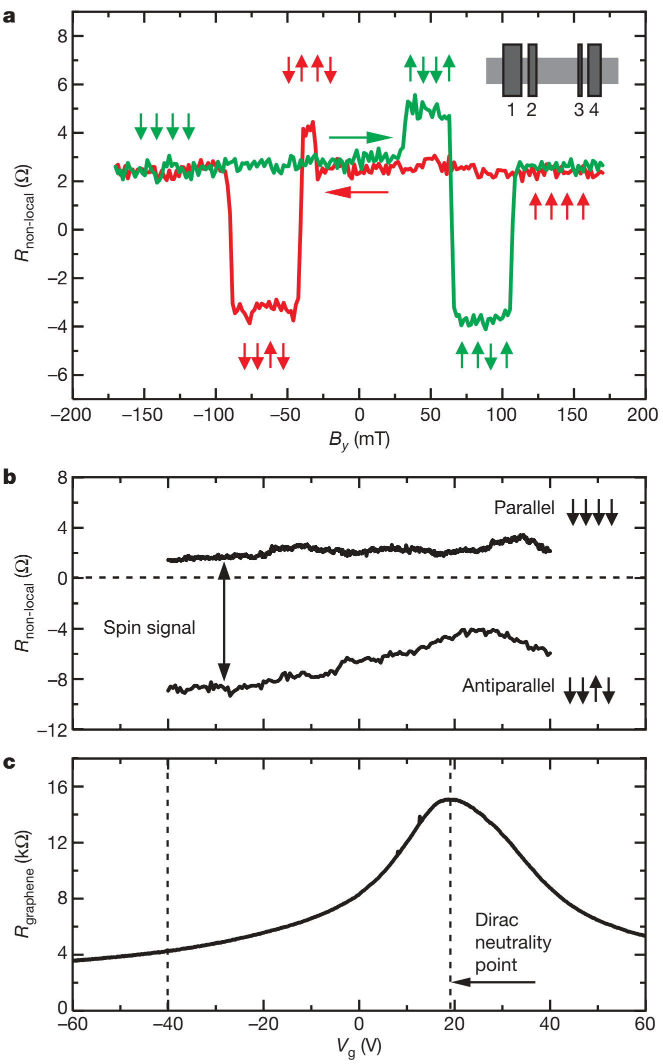


Figure 3 | Spin transport at room temperature. **a**, Non-local spin valve signal at room temperature for device 4. The magnetic field sweep directions are indicated as well as the magnetic configurations. The electrode widths are 500 nm (for electrode 1), 150 nm (for 2), 90 nm (for 3), 400 nm (for 4) and the electrode spacings were 450 nm for the gaps between electrodes 1 and 2, and between 3 and 4, and 3 μm for the gap between electrodes 2 and 3. The graphene width is 600 nm. Inset, electrode spacing configuration. **b**, Gate voltage dependence of the spin signals in parallel and anti-parallel directions. The spin signal has a weak minimum at the Dirac point. **c**, Gate voltage dependence of the graphene resistance. The maximum resistivity is 3.2 kΩ at the Dirac neutrality point ($V_g = 19$ V).

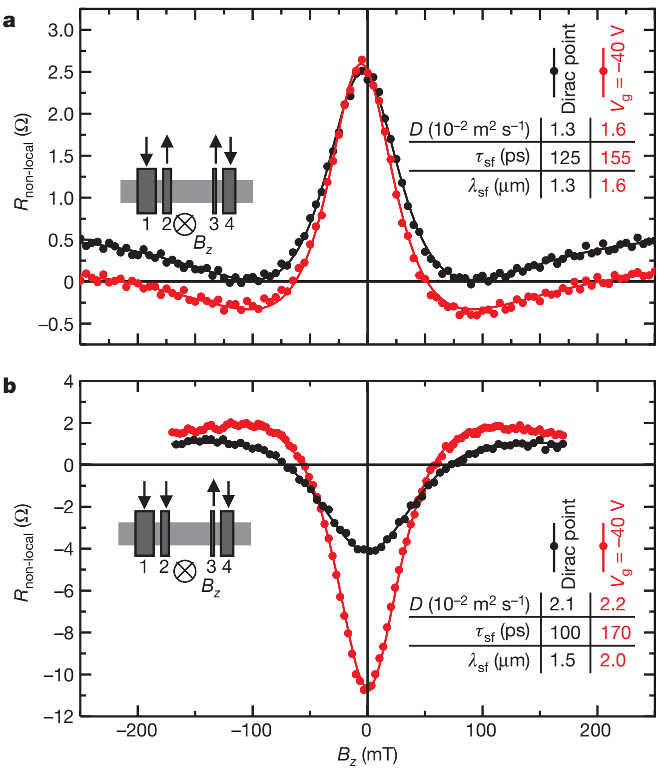


Figure 4 | Hanle spin precession. Hanle spin precession in the non-local geometry (device 4) measured as a function of the perpendicular magnetic field B_z for parallel (**a**) and antiparallel (**b**) configurations. By application of a magnetic field in the y direction the device is first prepared in the antiparallel/parallel configuration (see left insets). The data obtained at the Dirac point ($V_g = 19$ V) and at $V_g = -40$ V are compared (see tables in right insets). The solid lines are fits to the one-dimensional Bloch equations. The obtained parameters are shown in the insets, together with the corresponding spin relaxation lengths λ_{sf}.Note that the signal in **a** is much smaller than in **b**, owing to a sudden change in the properties of the 150-nm-wide contact between the measurements.

Here W is the width of the graphene strip, L the spacing of the central electrodes, and P is the spin polarization of the contacts. Inserting the obtained values of the measured spin signals at room temperature the formula gives $P \approx 0.1$.

Finally, we studied device 5 (graphene width, 1 μm) and found the spin signal to depend on the spacing L as follows: for $L = 1.5$ μm, the signal was 2.0 Ω; for 3 μm, 0.8 Ω; and for 5 μm, 0.25 Ω. These measurements are consistent with λ_{sf} in the range 1.5–2 μm and a longitudinal relaxation time T_1 of the same order as T_2, obtained previously.

In conclusion, we have demonstrated all electrical spin injection and detection in the two-dimensional carrier system formed by a single graphene layer. Spin transport is found to be relatively insensitive to the temperature. The use of the non-local technique together with artificial tunnel barriers has allowed us to observe spin transport. The contacts probably consist of relatively transparent regions that dominate the transport from the ferromagnet into the graphene. Optimization of these contacts is expected to result in even larger spin signals.

We believe that the observed spin relaxation length and relaxation times in our devices are limited by extrinsic impurity scattering and not by the intrinsic properties of graphene. The mobility of our devices is about 2,000 $cm^2 V^{-1} s^{-1}$. This is not exceptionally low[1], but improved fabrication techniques should allow for an increase of the scattering time, which will increase both D and τ_{sf}, and thus also λ_{sf}. This will make room-temperature spin transport possible over even longer distances, probably ultimately limited by electron–phonon scattering. At low temperatures the reduction of the role of spin–orbit interaction by combining high-mobility graphene layers with quantum confinement should make it possible to increase the spin relaxation times considerably.

METHODS SUMMARY

The devices (Fig. 1a) were prepared on a doped silicon substrate, insulated by a 500-nm-thick silicon oxide layer. A contact to the substrate allowed for a gate voltage to control the charge density in the graphene layer. By means of electron beam lithography, we deposited a Ti/Au marker pattern. Next, graphene flakes were deposited on the substrate using highly oriented pyrolytic graphite (HOPG, from Advanced Ceramics) by a repeated peeling technique[1]. The identification of the single graphene flakes is done by a combination of optical and atomic force microscopy. Sufficiently thin graphene flakes (<0.5 nm) and suitable sizes are selected and their positions relative to the markers are recorded. Then a 0.6-nm-thick layer of Al is deposited on the substrate under ultrahigh vacuum conditions. The substrate is cooled to 77 K, to increase the uniformity of the Al layer. The layer is oxidized under 100 mbar of O_2 pressure for 1 h and taken out to ambient atmosphere. Electrical measurements show that there is no conductance through this layer, indicating that the Al has been fully oxidized into Al_2O_3. Finally, using electron beam lithography the ferromagnetic contacts (50 nm thick) are fabricated. The devices are then either cooled down to 4.2 or 77 K (devices 1, 2 and 3) or measured at room temperature (devices 4 and 5), all under vacuum.

We measured contact resistances ranging between 5 and 50 kΩ, with a few above 100 kΩ. For example, the contact resistance of contact 2 is measured by passing current from contact 1 to 2, and measuring the voltage between 2 and 3. No clear scaling was found between the width of the contacts and their resistances, indicating some form of contact inhomogeneity. The gate voltage dependence of graphene on these devices is similar to that shown in Fig. 3c.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 March 2007; accepted 19 June 2007.
Published online 15 July 2007.

1. Geim, A. K. & Novoselov, K. S. The rise of graphene. *Nature Mater.* **6,** 183–191 (2007).
2. Heersche, H. B. *et al.* Bipolar supercurrent in graphene. *Nature* **446,** 56–59 (2007).
3. Trauzettel, B., Bulaev, D. V., Loss, D. & Burkard, G. Spin qubits in graphene quantum dots. *Nature Phys.* **3,** 192–196 (2007).
4. Huertas Hernando, D., Guinea, F. & Brataas, A. Spin orbit coupling in curved graphene, fullerenes, nanotubes and nanotube caps. *Phys. Rev. B* **74,** 155426 (2006).
5. Zutic, I. & Fuhrer, M. A path to spin logic. *Nature Phys.* **1,** 85–86 (2005).
6. Hill, E. W., Geim, A. K., Novoselov, K., Schedin, F. & Blake, P. Graphene spin valve devices. *IEEE Trans. Magn.* **42** (10), 2694 (2006).
7. Johnson, M. & Silsbee, R. H. Coupling of electronic charge and spin at a ferromagnetic paramagnetic metal interface. *Phys. Rev. B* **37,** 5312–5325 (1988).
8. Jedema, F. J., Heersche, H. B., Filip, A. T., Baselmans, J. J. A. & van Wees, B. J. Electrical detection of spin precession in a metallic mesoscopic spin valve. *Nature* **416,** 713–716 (2002).
9. Lou, X. *et al.* Electrical detection of spin transport in lateral ferromagnet-semiconductor devices. *Nature Phys.* **3,** 197–202 (2007).
10. Tombros, N., van der Molen, S. J. & van Wees, B. J. Separating spin and charge transport in single wall nanotubes. *Phys. Rev. B* **73,** 233403 (2006).
11. Schmidt, G. *et al.* Fundamental obstacle for spin injection from a ferromagnet into a diffusive semiconductor. *Phys. Rev. B* **62,** R4790–R4793 (2000).
12. Rashba, E. I. Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem. *Phys. Rev. B* **62,** R16267–R16270 (2000).
13. Fert, A. & Jaffres, H. Conditions for efficient spin injection from a ferromagnetic metal into a semiconductor. *Phys. Rev. B* **64,** 184420 (2001).
14. Costache, M. V., Zaffalon, M. & van Wees, B. J. Spin accumulation probed in multiterminal lateral all-metallic devices. *Phys. Rev. B* **74,** 012412 (2006).
15. Jedema, F. J., Nijboer, M. S., Filip, A. T. & van Wees, B. J. Spin injection and spin accumulation in all-metal mesoscopic spin valves. *Phys. Rev. B* **67,** 085319 (2003).
16. Ando, T. Screening effect and impurity scattering in monolayer graphene. *J. Phys. Soc. Jpn* **75,** 074716 (2006).
17. Peres, N. M. P. *et al.* Coulomb interactions and ferromagnetism in pure and undoped graphene. *Phys. Rev. B* **72,** 174406 (2005).

Acknowledgements We thank B. Wolfs, S. Bakker, E. Koop, A. Morpurgo, R. Pandian, G. Palasantzas and N. Katsonis for technological support and discussions. We acknowledge financial support from the NWO (PIONIER grant) and MSC^{plus}.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.T. (n.tombros@rug.nl).

METHODS

Measurements on devices with transparent low ohmic contacts. Here we give a brief overview of devices with transparent and low ohmic contacts. The preparation is as described in the main text except for the Al evaporation and oxidation step. In two of these devices the ferromagnetic electrodes did not fully overlap the graphene flakes. As a result the graphene is exposed to stray magnetic fields in the range of several hundreds of milliTesla at the ends of the strips. In four terminal measurements at 4.2 K we observed magnetoresistances of the order of 300 Ω on a background of 18 kΩ. These resistances coincided with the expected switching fields of the cobalt electrodes. Moreover, the sign of the signals could change as function of the gate voltage but also as a function of the bias voltage. The sensitivity to magnetic field, charge density and energy strongly suggest that these are Hall-type effects and we conclude that to avoid stray fields, the ferromagnets should completely overlap the graphene.

A third device had a geometry similar to device 1. Here we found three-terminal contact resistances in the range $<50\,\Omega$ (hence close to zero). This implies that there is a good contact between the ferromagnet and the graphene underneath, and the carriers are not able to pass underneath the contacts. This was confirmed by measurements of the non-local resistance, which yielded low values ($<1\,\Omega$). No spin valve signals could be detected. We also measured the two terminal local spin valve signals between contacts 2 and 3. On a background of typically 5 kΩ no signals could be observed, the noise level being about 5 Ω.

Spin relaxation induced by the contacts. We found the lowest contact resistances (devices 1, 2, 3, 4 and 5) to be around 5 kΩ, which corresponds to a resistance of about 10 kΩ per spin channel. This implies that a contact connects the spin-up and spin-down channels in the graphene by an effective resistance of 20 kΩ. Taking into account all four contacts, an effective relaxation time can thus be calculated from the spin relaxation via the contacts (see ref. 18). An estimate shows that this time exceeds the observed spin relaxation time by at least a factor of ten. The contacts therefore contribute at most 10% to the observed spin relaxation.

Experimental aspects. The slight asymmetry in coercive fields of the ferromagnetic strips for positive and negative B sweeps arises from a 'frozen-in' magnetization configuration in the wider parts of the ferromagnetic leads. Full symmetry can be restored by continuing the B sweeps to higher magnetic fields ($\sim$1 T). In most cases this was not done for practical reasons.

Non-idealities in the non-local resistance measurements (in particular related to nonhomogeneous contacts) can generate a spin-independent background. This is estimated to be less than 1 Ω in typical measurements.

18. Zaffalon, M. & van Wees, B. J. Zero dimensional spin accumulation in a mesoscopic island. *Phys. Rev. Lett.* **91,** 186601 (2003).

LETTERS

Warming trends in Asia amplified by brown cloud solar absorption

Veerabhadran Ramanathan[1], Muvva V. Ramana[1], Gregory Roberts[1], Dohyeong Kim[1], Craig Corrigan[1], Chul Chung[1] & David Winker[2]

Atmospheric brown clouds are mostly the result of biomass burning and fossil fuel consumption[1]. They consist of a mixture of light-absorbing and light-scattering aerosols[1] and therefore contribute to atmospheric solar heating and surface cooling. The sum of the two climate forcing terms—the net aerosol forcing effect—is thought to be negative and may have masked as much as half of the global warming attributed to the recent rapid rise in greenhouse gases[2]. There is, however, at least a fourfold uncertainty[2] in the aerosol forcing effect. Atmospheric solar heating is a significant source of the uncertainty, because current estimates are largely derived from model studies. Here we use three lightweight unmanned aerial vehicles that were vertically stacked between 0.5 and 3 km over the polluted Indian Ocean. These unmanned aerial vehicles deployed miniaturized instruments measuring aerosol concentrations, soot amount and solar fluxes. During 18 flight missions the three unmanned aerial vehicles were flown with a horizontal separation of tens of metres or less and a temporal separation of less than ten seconds, which made it possible to measure the atmospheric solar heating rates directly. We found that atmospheric brown clouds enhanced lower atmospheric solar heating by about 50 per cent. Our general circulation model simulations, which take into account the recently observed widespread occurrence of vertically extended atmospheric brown clouds over the Indian Ocean and Asia[3], suggest that atmospheric brown clouds contribute as much as the recent increase in anthropogenic greenhouse gases to regional lower atmospheric warming trends. We propose that the combined warming trend of 0.25 K per decade may be sufficient to account for the observed retreat of the Himalayan glaciers[4–6].

Field experiments[1,7] during the past decade over the polluted oceans provided observational constraints for the regional aerosol forcing, but these studies were limited to the vertically integrated forcing of the surface–atmosphere system. Crucial to our understanding of the tropical regions would be direct observations of the vertical heating rate. During the six-month-long tropical dry season, convective coupling between the surface and the troposphere is weak. As a result, aerosol solar heating can amplify the effect of greenhouse gases in warming the atmosphere while simultaneously cooling the surface[1]. The primary motivation for the Maldives Autonomous unmanned aerial vehicle Campaign (MAC), the results of which are reported here, was the observational determination of this solar heating and the examination of its role in atmospheric warming trends.

MAC extends the pioneering lightweight unmanned aerial vehicle (UAV) campaigns initiated by other investigators[8] by stacking three UAVs and synchronizing the heating rate measurements with measurements of aerosols, clouds, humidity and temperature. The downward and reflected broadband (305 to 2,800 nm) and visible (400 to 700 nm) solar fluxes were measured typically at 0.5, 1.5 and 3 km. The net fluxes (down minus up) were differenced to obtain the heating rate for the layer between the flight levels.

MAC was conducted over the Indian Ocean from the island of Hanimadhoo (6.78° N, 73.18° E) during March 2006. The UAV measurements were validated with identical measurements at the Maldives Climate Observatory at Hanimadhoo (MCOH)[9] during low-altitude overflights. March is part of the dry season (November to May), when low-level flow brings polluted air-mass from Asia to most of the northern Indian Ocean[1]. During 4–16 March (period 1), the air-mass was of marine origin, but was still influenced by atmospheric brown clouds (ABCs) with near-surface aerosol number concentrations as high as 1,000 cm^{-3}. The aerosol layer was confined to the boundary layer (Fig. 1a). During 19–29 March (period 2), MAC detected 1–3-day-old air mass from southern Asia above 1 km (Supplementary Fig. 1). The measured aerosol absorption coefficient above 1 km increased tenfold (Fig. 1a), equivalent to a 700 ng m^{-3} increase of black carbon mass. The aerosol number concentration increased from about 750 to about 2,500 cm^{-3}. The boundary-layer aerosol number concentration was similar to that during period 1. The measured visible aerosol optical depth, AOD, a good index for column-averaged scattering and absorption, was 0.18 during period 1 and 0.4 during period 2. The visible absorption optical depth, AOD_{abs}, increased from about 0.005 to 0.027 from periods 1 to 2.

The observed heating rates are shown in Fig. 1b and c together with the simulated values using the Monte Carlo Aerosol Cloud Radiation model (MACR) described elsewhere[1]. MACR used MAC and MCOH data for aerosol and cloud parameters, H_2O and O_3. For the 0.5–3 km layer, the mean period 1 diurnal average heating rates H for MAC and MACR were respectively 1.1(±0.3) and 1.2(±0.3) K per day; the period 2 H values were respectively 1.8(±0.2) and 1.65(±0.2) K per day. The agreement between MAC and MACR demonstrates that it is possible to measure H accurately in the atmosphere using multiple aircraft and link the H value with observations of atmospheric constituents using models. Figure 1d shows the magnitude of the heating that is due to the addition of ABCs from southern Asia: ΔH, the difference in H over 0.5–3 km between the two periods. The ΔH from MACR is about 0.45(±0.3) K per day compared with the MAC value of 0.7(±0.3) K per day. The difference is within experimental and modelling uncertainties. MACR simulations show that about 90% of the ΔH is attributable to soot. The balance is due to small changes in water vapour and clouds. The major role of soot is also confirmed by the increase in heating of about 0.25(±0.15) K per day detected by the visible radiometer. Elemental and organic carbon in soot are the dominant absorbers[10,11] in the visible wavelengths. In

[1]Center for Clouds, Chemistry and Climate, Scripps Institution of Oceanography, UCSD, La Jolla, California 92037, USA. [2]NASA Langley Research Center, Hampton, Virginia 23681-0001, USA.

addition, direct measurements of elemental carbon mass from filter samples at MCOH correlated strongly[12] with the measured aerosol absorption coefficient. Our main deduction from Fig. 1d is that ABCs from southern Asia increased AOD and AOD_{abs} respectively by 0.22 and 0.02, which was sufficient to enhance H by about 50%. This plume is part of the widespread ABC that envelops most of Asia and the adjacent oceans, as documented extensively[1,13,14]. Hence the UAV-measured large solar heating rates have climatic implications, which we consider below.

In what follows, we will rely on the estimates of Chung *et al.*[14], because this study used both satellite and ground-based observations to account for the spatial and temporal variations of ABCs. AODs, solar heating and surface dimming were obtained for the period from 2000 to 2003 (ref. 14). The monthly AOD (and AOD_{abs}) values over India varied from 0.1 (and 0.01) in July to 0.3 (and 0.03) during January, whereas over China they increased from 0.1 (and 0.01) in January to 0.35 (and 0.04) in July (Supplementary Fig. 2). Comparing these with the MAC AOD values and ΔH, we conclude that ABCs can enhance the solar heating rates substantially on spatio-temporal scales relevant for climate impact studies. The impact on climate depends also on the vertical extent of the plume[1], but such data were unavailable until recently.

The vertical extent data are now available through the Cloud Aerosol Lidar and Infrared Pathfinder Satellite (CALIPSO)[3], launched on April 2006. CALIPSO reveals that vertically extended ABCs are widespread through most of southern and eastern Asia (Fig. 2) with peak lidar (light detection and ranging) returns of between 1 to 3 km above sea level (a.s.l.), which is indicative of aerosol concentrations and is similar to the MAC vertical extent (Fig. 1a). The 1–3 km a.s.l. plume is persistent from November 2006 to March 2007 (Fig. 2a–d and Supplementary Fig. 3) over the region of interest.

Using aircraft data over the Indian Ocean[1], Chung *et al.*[14] assumed that ABCs extended to 3 km a.s.l. over Asia, and this assumption is now supported by the CALIPSO data. Because the AOD and AOD_{abs} values were also constrained by observations, their forcing values provide a sound starting point for evaluating the climate impacts. For example, the 0–3 km layer heating per unit AOD_{abs} from MAC (Fig. 1d) is about 32 K per day per unit AOD_{abs} compared with the ref. 14 value of 23 K per day per unit AOD_{abs}. We inserted monthly mean ABC solar heating rates[14] (Supplementary Fig. 4) and dimming in the CCM3 version[15] of the NCAR general circulation model (GCM) and conducted simulations (see Methods).

The tropical average model atmosphere warms in response to the imposed ABC heating, with peak annual mean warming of 0.4 K between 2 and 4 km a.s.l. (Supplementary Fig. 5). The annual warming at 700 mb (about 3 km a.s.l.) over the Himalayan-Hindu-Kush region is about 0.6 K (Fig. 3), with peak warming during March to September (Supplementary Fig. 6). Although ABC heating is restricted to the lower troposphere (pressure $P > 700$ mb), the warming penetrates well into the mid-troposphere (500 mb) owing to vertical transport (Supplementary Fig. 7).

As reported in ref. 16, emissions of the ABC precursors SO_2 and soot increased tenfold from 1930 to 2000. Using these emissions data, it has been shown[13] that more than 90% of the simulated ABC aerosol forcing and the induced atmospheric warming is attributable to the period from 1950 to present. We obtained the lower atmosphere warming simulated by the NCAR model for the greenhouse gas forcing for 1950 to 2000 (ref. 17). For the 2 to 5 km (a.s.l.) layer, this warming is about 0.5–0.8 K. The ratio of the ABC warming and the 1950–2000 greenhouse gas warming is about 1 over the Himalayan-Hindu-Kush region (Supplementary Fig. 7). The combined simulated warming of greenhouse gases and ABCs at the higher levels (3–5 km a.s.l.) is about 1.2 K, or about 0.25 K per decade during

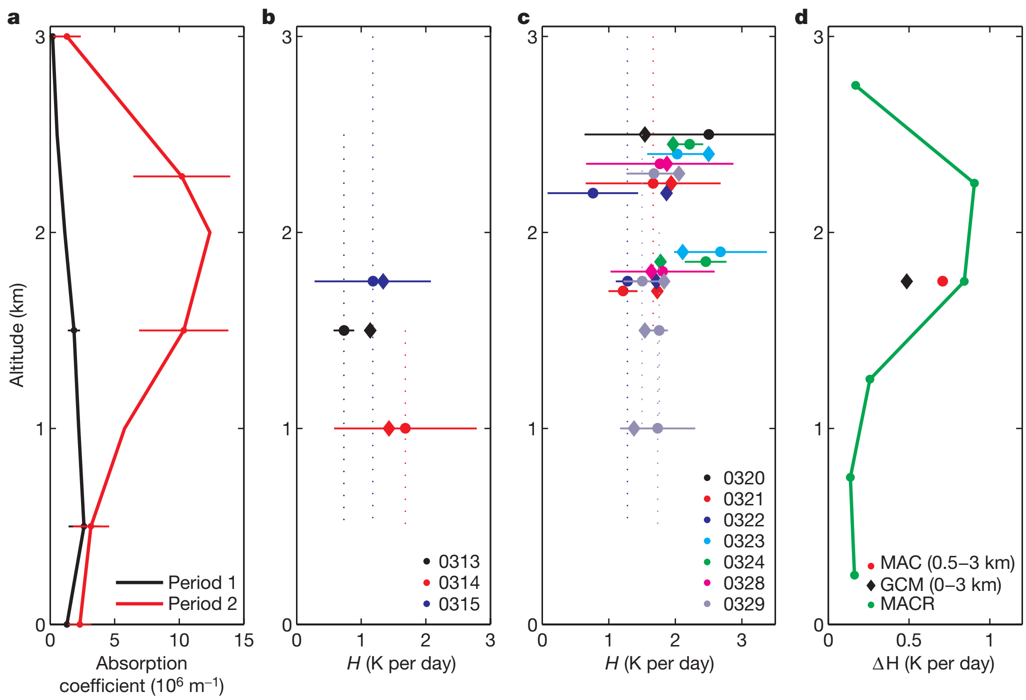


Figure 1 | Vertical distribution of UAV observations. The MAC period is divided into two periods. Period 1 covers 4–16 March, when aerosol was mainly below 1 km; period 2 covers 19–29 March, when aerosol was elevated. **a**, Aerosol absorption coefficient. **b**, Diurnally averaged solar heating rate H for period 1. The instantaneous fluxes were normalized to diurnal average solar fluxes using MACR simulations to obtain diurnal mean H. The vertical bar shows the layer depth; the horizontal bar shows the variations between flight legs, with each flight lasting about 25 min; solid circles are observed values and diamonds are simulated values by MACR. The observation dates are shown (for example 0313 denotes 13 March). **c**, As for **b** but for period 2. **d**, Difference in the vertically (0.5–3 km) averaged H between period 2 and period 1. The green line is the MACR profile for ΔH. The spatial average over the Indian Ocean and southern Asia of the specified CCM3 ABC heating is also shown (GCM). The uncertainty in H is 0.3 K per day for period 1 and 0.2 K per day for period 2. All uncertainty estimates, including those in the text, are 2σ values.

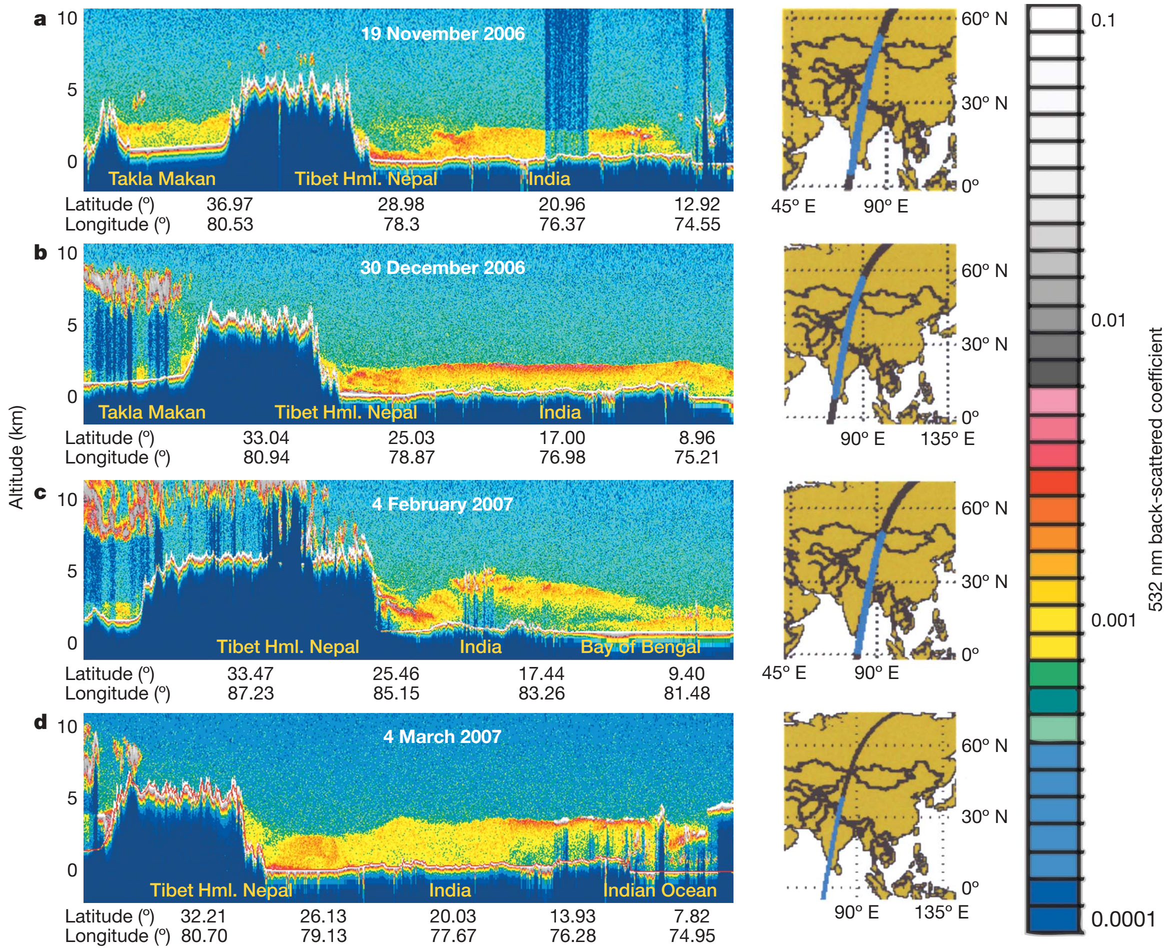


Figure 2 | Colour-coded profiles of 532 nm backscatter return signal from the CALIPSO lidar showing the vertical distribution of ABCs. The panels on the right show the orbit track across Asia and on the left the vertical extent of the aerosol is shown for the blue-shaded portion of each track. We chose the colour scale so that aerosol usually shows up in green, yellow and red (for low, medium and high loadings, respectively) and boundary layer clouds usually show up as grey or white. Cirrus usually ranges from yellow to grey. Sample profiles are shown for four months of the dry season that extends from November to May. The Takla Makan desert is in northwestern China between 37° N and 41° N and 77° E to 90° E. Hml., Himalayas.

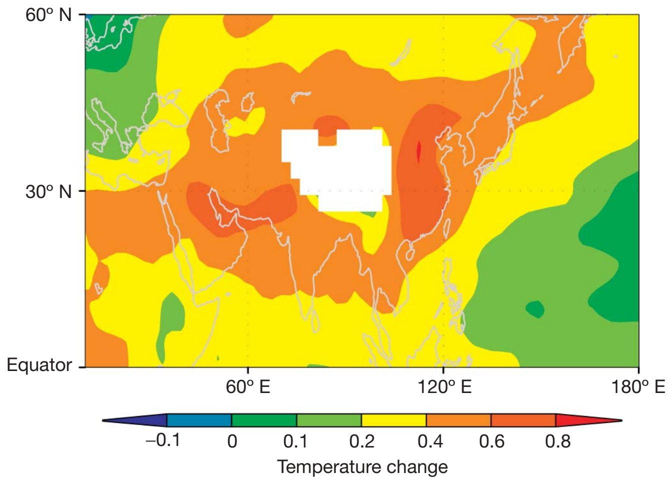


Figure 3 | The temperature change d*T* due to ABC solar heating at the 700 hPa (or 700 mb) level (~3 km a.s.l.) simulated by CCM3. The change is estimated from a 50-year average of experiment and control. The standard error of the mean difference is 0.05 K. All altitudes mentioned in the text are a.s.l. altitudes. For example, a warming of 0.6 K at 3 km a.s.l. would denote near-surface warming for a location in the Himalayas that is at an elevation of 3 km a.s.l. The white-shaded regions are the part of the Himalayas with surface pressure less than 700 mb, that is, surface elevations are higher than 3 km.

the period from 1950 to present, which is about twice the simulated surface warming trend (Supplementary Figs 5 and 6). This result is also supported by satellite microwave temperature trends from 1979 to 2000 that reveal that the atmospheric trend is larger than the surface trend by about 0.15 K per decade[13] (Supplementary Fig. 8).

The above model results have significant implications because the observed air-temperature trend over the elevated Himalayas has accelerated to between 0.15–0.3 K per decade[18] during the past several decades. This large trend is thought to be the major cause for the Himalayan-Hindu-Kush glacier ablation[5]. The Himalayan-Hindu-Kush region has seen a marked retreat in the glaciers that serve major Asian rivers such as the Yangtze, the Indus and the Ganges[4,19]. The rapid melting of these glaciers, the third-largest ice mass on the planet, if it becomes widespread and continues for several more decades, will have unprecedented downstream effects on southern and eastern Asia[4,5].

METHODS SUMMARY

MAC. The MAC campaign, platforms and instruments are described in ref. 20. The miniaturized instruments are described in ref. 21. The radiation and aerosol instrument measurements were validated with MCOH measurements[22,23]. The overall MAC observations have also been presented[24]. The cloud fraction, measured from the in-cloud UAV cloud droplet instrument, was 11% for period 1 and 5% for period 2. The water vapour amount from 0 to 3 km was 35.1 kg m^{-2} for period 1 and 34.3 kg m^{-2} for period 2.

MACR. Sensitivity studies with MACR were used to understand the factors contributing to ΔH. Because changes in cloud fraction and water vapour amount

between the two periods were small (see above), the net effect of these two parameters on ΔH was only $+0.05$ K per day. About 10–12% of ΔH was in wavelengths $\lambda < 400$ nm; 37% to 44% for $400 < \lambda < 700$ nm and 45% to 54% for $\lambda > 700$ nm. Soot contributed 100% of the increased absorption for $\lambda < 700$nm and about 70% for $\lambda > 700$ nm.

CCM3. The CCM3 is run with prescribed sea surface temperatures. Chung *et al.*[14] distributed the AOD uniformly from the surface to about 3 km. The solar heating would have been larger by about 50% had we adopted an elevated plume (as in Fig. 1a) instead of the uniform aerosol profile. To compare the ABC warming with that due to the observed warming of the sea surface, we prescribed the observed sea surface temperature trend for the 1950 to 2002 period for each grid point in the model as a function of the month, and ran the model with and without the prescribed aerosol forcing. Four ensemble runs of 52 years each were conducted for the control and the experiments.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 February; accepted 13 June 2007.

1. Ramanathan, V. *et al.* Indian Ocean experiment: An integrated analysis of the climate forcing and effects of the great Indo-Asian haze. *J. Geophys. Res.* **106,** doi:10.1029/2001JD900133 (2001).
2. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis (Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change)* 131–217 (World Meteorological Organization, Geneva, 2007).
3. Winker, D. M., Vaughan, M. A. & Hunt, W. H. The CALIPSO mission and initial results from CALIOP. In *Lidar Remote Sensing for Environmental Monitoring VII*; *Proc. SPIE* **6409,** 1–8 (2006).
4. Barnett, T. P., Adam, J. C. & Lettenmaier, D. P. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **438,** 303–309 (2005).
5. Thompson, L. G. *et al.* Tropical glacier and ice core evidence of climate changes on annual to millennial time scales. *Clim. Change* **59,** 137–155 (2003).
6. Vuille, M., Bradley, R. S., Werner, M. & Keimig, F. 20th century climate change in the tropical Andes. *Clim. Change* **59,** 75–99 (2003).
7. Haywood, J. *et al.* Radiative properties and direct radiative effect of Saharan dust measured by the C-130 aircraft during SHADE: 1. Solar spectrum. *J. Geophys. Res.* **108,** doi:10.1029/2002JD002687 (2003).
8. Holland, G. H. *et al.* The aerosonde robotic aircraft: A new paradigm for environmental observations. *Bull. Am. Meteorol. Soc.* **82,** 889–901 (2001).
9. Ramana, M. V. & Ramanathan, V. Abrupt transition from natural to anthropogenic aerosol radiative forcing: Observations at the ABC-Maldives Climate Observatory. *J. Geophys. Res.* **111,** doi:10.1029/2006JD007063 (2006).
10. Andrea, M. O. & Gelencser, A. Black carbon or brown carbon? The nature of light absorbing carbonaceous aerosols. *Atmos. Chem. Phys* **6,** 3131–3148 (2006).
11. Ramaswamy, V. *et al.* in *Climate Change 2001: The Scientific Basis (Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change)* (Cambridge Univ. Press, Cambridge, UK, 2001).
12. Corrigan, C. E., Ramanathan, V. & Schauer, J. J. Impact of monsoon transitions on the physical and optical properties of aerosols. *J. Geophys. Res.* **111,** doi:10.1029/2005JD006370 (2006).
13. Ramanathan, V. *et al.* Atmospheric brown clouds: impacts on south Asian climate and hydrological cycle. *Proc. Natl Acad. Sci. USA* **102,** 5326–5333 (2005).
14. Chung, C. E., Ramanathan, V., Kim, D. & Podgorny, I. A. Global anthropogenic aerosol direct forcing derived from satellite and ground-based observations. *J. Geophys. Res.* **110,** doi:10.1029/2005JD006356 (2005).
15. Kiehl, J. T. *et al.* The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Clim.* **11,** 1131–1149 (1998).
16. Novakov, T. *et al.* Large historical changes of fossil-fuel black carbon aerosols. *Geophys. Res. Lett.* **30,** doi:10.1029/2002GL016345 (2003).
17. Washington, W. M. *et al.* Parallel climate model (PCM) control and transient simulations. *Clim. Dyn.* **16,** 755–774 (2000).
18. Liu, X. D. & Chen, B. D. Climatic warming in the Tibetan plateau during recent decades. *Int. J. Climatol.* **20,** 1729–1742 (2000).
19. Kulkarni, A. V., Rathore, B. P., Mahajan, S. & Mathur, P. Alarming retreat of Parbati glacier, Beas basin, Himachal Pradesh. *Curr. Sci.* **88,** 1844–1850 (2005).
20. Ramanathan, V., Roberts, G., Corrigan, C., Ramana, M. V. & Nguyen, H. *Maldives AUAV Campaign (MAC): Observing Aerosol-Cloud-Radiation Interactions Simultaneously from Three Stacked Autonomous Unmanned Aerial Vehicles (AUAVs).* <http://www-abc-asia.ucsd.edu/MAC/MAC_proposal_FINAL_2005July05.pdf> (2005).
21. Roberts, G. C., Ramanathan, V., Corrigan, C. & Ramana, M. V. Aerosol, cloud, and radiometric measurements with small autonomous unmanned aerial vehicles. *Am. Assoc. Aerosol Res. Conf. (Austin, Texas, 2005)* ⟨http://www-abc-asia.ucsd.edu/MAC/RobertsAAARC2005.pdf⟩ (2005).
22. Ramana, M. V. *et al.* Direct measurements of albedo and solar absorption over the Northern Indian Ocean with a new observing system of stacked multiple AUAVs. *Eos* **87** (52), abstr. A13B–0921 (2006).
23. Corrigan, C. E., Ramanathan, V., Ramana, M. V., Kim, D. & Roberts, G. Vertical profiles of aerosols using unmanned aerial vehicles. *Int. Aerosol Conf.* (St Paul, Minnesota, 2006) ⟨http://www-abc-asia.ucsd.edu/MAC/Corrigan_IAC_20062.pdf⟩ (2006).
24. Ramanathan, V. *et al.* Simultaneous observations of aerosols, clouds, and radiometric fluxes using light-weight autonomous UAVS. *Eos* **87** (52), abstr. A33B–1007 (2006) <http://www-abc-asia.ucsd.edu/MAC/RamanathanAGU2006SimultaneousAbstract.pdf>.
25. Dubovik, O. *et al.* Accuracy assessment of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) sun and sky radiance measurements. *J. Geophys. Res.* **105,** 9791–9806 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. Three summary figures are also included.

Acknowledgements We acknowledge NSF and J. Fein for supporting and guiding the UAV programme and the NSF, the NOAA and NASA for funding the MAC campaign. We also thank D. Fahey, H. Maring, J. Kuettner, C. Jennison and R. Curry for guidance of MAC and Advanced Ceramics for the field support and the government of Maldives for support of the field campaign and MCOH.

Author Contributions V.R. designed MAC, performed the analysis and wrote the paper. M.V.R. was responsible for radiometric observations, validation and data analysis. G.R. was the lead instrument scientist; designed the instrument–UAV integration package; and did the analysis of aerosol and cloud physics data. D.K. conducted the MACR calculations and analysis. C. Corrigan was responsible for UAV absorption measurements and analysis of this data. C. Chung was responsible for GCM simulations and analysis. D.W. is the Principal Investigator for CALIPSO data. All authors reviewed and commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.R. (vramanathan@ucsd.edu).

METHODS

UAVs. The UAV is the Manta aircraft built by Advanced Ceramics Research (Arizona, USA). The below-cloud UAV flew below the trade cumulus cloud base at about 0.5 km and the above-cloud UAV flew mostly at 3 km; both UAVs measured boundary-layer aerosol number concentration, size distribution and absorption coefficient at 370, 520 and 880 nm to determine concentration of absorbing aerosols. Up- and down-looking broadband (305 to 2,800 nm) pyranometers (Kipp & Zonen) and visible-band (400–700 nm) radiometers (LICOR) measured the transmitted (from above) and reflected (from below) solar fluxes. The in-cloud UAV flying through the middle of the cloud, typically at about 750 m to 1 km, measured cloud drop number size distribution from 2 to 50 μm in diameter. The instruments and the power supply together weighed 3.5–5.4 kg. When the wind vertical shear was less than 25 m s^{-1}, the UAVs were able to maintain their altitudes and their designed vertical and horizontal separation.

Validation with MCOH. This ground-based observatory[9,12] has been collecting data since 2004. It contains aerosol and radiation instruments identical to those on the UAVs and each flight included a low-level flight within 5 km of MCOH. The aerosol number concentration from the UAVs (below 0.5 km) and the absorption coefficient agreed with MCOH observations within about 10% (ref. 23). The radiometers were validated by two independent methods. We flew two UAVs side by side: the radiometers in the two UAVs agreed within 5 W m^{-2} (0.5%). Next, we conducted fly-bys over the MCOH and the UAV radiometer agreed with the MCOH radiometers[22] within 5 W m^{-2}. The AOD and the AOD_{abs} at 520 nm were measured by the AERONET–CIMEL Sun photometer[25].

MACR simulations. The input to MACR was AERONET data for AOD, AOD_{abs}, and columnar water vapour. AODs at four wavelengths from 440 to 1,020 nm were obtained from AERONET and then extrapolated to all other wavelengths using the 'Angstrom exponent'[9]. Within each layer from 0 to 3 km, we used the measured vertical distribution of aerosol concentration to scale the column AODs to AODs for each layer. For validation of this procedure, the MACR values were compared with the measured absorption coefficient at $\lambda = 370$, 520 and 880 nm: the measured and MACR values were within 10% of each other. The columnar ozone was measured using a hand-held multi-spectral Sun photometer. The procedure for estimating uncertainty in the measured and simulated heating rates is explained in the Supplementary Information.

CALIPSO. This satellite carries the first polarization lidar in space with lasers operating at 532 and 1,064 nm (ref. 3). The polarization enables us to distinguish between spherical cloud particles and non-spherical aerosols such as dust and soot. In addition, the dual wavelengths are used to retrieve aerosol size.

CCM3 input for the ABC forcing. The Chung *et al.* aerosol forcing was used in the CCM3 simulations. They[14] used recent aerosol satellite data from the MODIS instrument on NASA's TERRA satellite and ground-based aerosol robotics network[25] as the basis for the aerosol optical depths and single scattering albedo; and then used a combination of NASA's aerosol assimilation model and small mode aerosol radii from MODIS to determine the anthropogenic fraction of the observed AODs and single scattering albedos. Using this global distribution of anthropogenic AODs and single scattering albedo, Chung *et al.* introduce this data in the MACR model with satellite climatology of clouds to derive reduction of solar radiation at the surface and the increase in atmospheric solar absorption due to ABCs for the period 2000–2003. These estimates[14] were restricted to 2000–2003 because the MODIS data were available for this period, as were the GOCART values for the anthropogenic fraction. But they assumed that the aerosol plume was contained between the surface to 3 km a.s.l.; this was based on the Indian Ocean Experiment[1] observations over the Arabian sea.

Derived trends for the 1950–2000 period. The ABC-induced temperature changes shown in this study were estimated by comparing the CCM3 simulations that used the imposed 2000–2003 forcing with the simulations that did not use the ABC forcing. We then use ref. 13 to conclude that the computed warming must have occurred during the period 1950–2000. This coupled ocean–atmosphere model study[13], restricted to southern Asia, used the emission data of ABC precursors (SO_2 and soot) from 1930 to 2002. They took the present-day forcing from observations and aerosol assimilation models (this procedure was followed by ref. 14) for the period 1995 to 1999; and scaled this forcing with the published data for soot and SO_2 emission rates (as a function of time) to derive a time-dependent forcing from 1930 to 2000. Their emission data and the derived atmospheric solar heating and dimming show that more than 90% of the forcing occurred after 1950 (figure 1 in ref. 13). Ref. 14 validated the simulated trend in dimming by comparing the trend with the observed solar radiation trends over India (see top panel of figure 2 in ref. 13).

Vol 448 | 2 August 2007 | doi:10.1038/nature06017

LETTERS

Non-volcanic tremor driven by large transient shear stresses

Justin L. Rubinstein[1], John E. Vidale[1], Joan Gomberg[2], Paul Bodin[1], Kenneth C. Creager[1] & Stephen D. Malone[1]

Non-impulsive seismic radiation or 'tremor' has long been observed at volcanoes[1] and more recently around subduction zones[2]. Although the number of observations of non-volcanic tremor is steadily increasing, the causative mechanism remains unclear. Some have attributed non-volcanic tremor to the movement of fluids[2–6], while its coincidence with geodetically observed slow-slip events at regular intervals[7,8] has led others to consider slip on the plate interface as its cause[7–14]. Low-frequency earthquakes in Japan, which are believed to make up at least part of non-volcanic tremor[9], have focal mechanisms[10] and locations[11] that are consistent with tremor being generated by shear slip on the subduction interface. In Cascadia, however, tremor locations appear to be more distributed in depth than in Japan[3,4], making them harder to reconcile with a plate interface shear-slip model. Here we identify bursts of tremor that radiated from the Cascadia subduction zone near Vancouver Island, Canada, during the strongest shaking from the moment magnitude $M_w = 7.8$, 2002 Denali, Alaska, earthquake. Tremor occurs when the Love wave displacements are to the southwest (the direction of plate convergence of the overriding plate), implying that the Love waves trigger the tremor. We show that these displacements correspond to shear stresses of approximately 40 kPa on the plate interface, which suggests that the effective stress on the plate interface is very low. These observations indicate that tremor and possibly slow slip can be instantaneously induced by shear stress increases on the subduction interface—effectively a frictional failure response to the driving stress.

To understand the source physics of non-volcanic tremor we need to know how it initiates, and more specifically whether it can be initiated by dynamic stresses. Obara[2] identified three different earthquakes of local magnitude $M_J \geq 5$ that may have triggered tremor. Following this initial study, Obara[15] identified many more earthquakes that may have triggered tremor, including local micro-earthquakes ($2 \leq M_J \leq 3$), local medium magnitude earthquakes ($3 \leq M_J \leq 6.5$), and large earthquakes ($M_w \geq 7.0$ at $>$1,000 km).

These findings led Miyazawa and Mori to closely examine tremor triggered by the $M_w = 8.3$ Tokachi-Oki[16] and the 2004 Sumatra earthquakes[17]. For the specific geometry of the Sumatra earthquake and its orientation relative to the Japanese subduction zone, dilatational stresses associated with the Rayleigh wave were inferred to be much more effective than horizontal shear stresses in triggering tremor[17]. On the basis of this finding, it was hypothesized that fluid flow facilitates tremor[17,18].

Inspired by the above findings, we examined seismic recordings of the 3 November 2002 $M_w = 7.8$ Denali earthquake, which triggered seismicity throughout much of western North America[19]. In this case, directivity focused and amplified seismic waves towards western North America, especially the Love waves[19,20]. Examining these seismograms, we found clear evidence of triggered tremor near Vancouver Island.

The tectonics of the region are dominated by the slow subduction of the Juan de Fuca and Explorer plates underneath the North American plate, with the trench located off the west coast of Vancouver Island[21] (Fig. 1). This subduction zone is well known because it is the place where episodic tremor and slip (ETS), the repeated coincidence of geodetically observed slow-slip and elevated levels of non-volcanic tremor at regular intervals, was first discovered[7]. Non-volcanic tremor has been observed running almost the entire length of Vancouver Island[3,22]. Examining the seismic records of the Denali earthquake in Vancouver Island, we identified shaking-induced tremor bursts in northern Vancouver Island. There are indications that tremor was also triggered on Southern Vancouver Island, but it was overprinted by stronger, triggered earthquake activity. Thus, we focused on the northern region.

In the recordings of the Denali mainshock at high frequencies we can identify Denali's P waves, locally triggered tremor and locally triggered earthquakes (Fig. 2). At low frequencies the Love and Rayleigh waves from Denali are visible. The tremor appeared in five bursts that peaked at 310, 345, 365, 386 and 415 s at station PHC, with distinct gaps between some of the bursts. There are additional patterns of high-frequency energy that differed for each station, especially in the interval from 425 to 500 s, whose origin is not clear. The

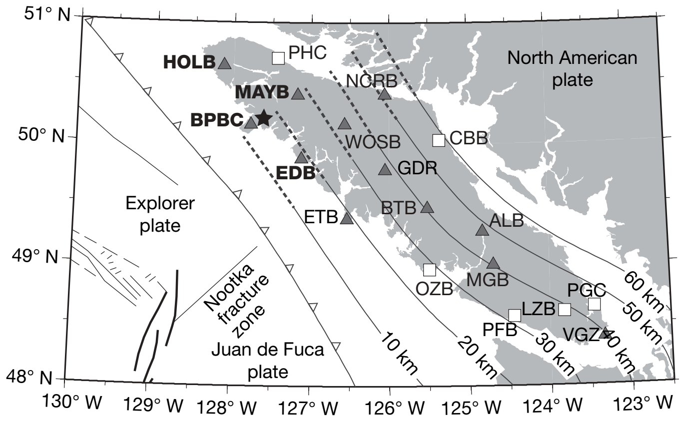


Figure 1 | Map of Vancouver Island. Local geologic features, Canadian National Seismic Network (CNSN) stations on the island (white squares, broadband; grey triangles, short period), and the location of the triggered tremor (black star) are indicated on the map. The stations we used to locate the triggered tremor are indicated with bold type. Solid lines indicate isodepths (shown in kilometres) of the interface between the Juan de Fuca and the North American plates; dotted lines indicate isodepths of the interface between the Explorer and the North American plates; solid line with triangles indicates the location of the plate interface at the earth's surface. Figure modified with permission from figure 1b in ref. 22.

[1]Department of Earth and Space Science, [2]United States Geological Survey, University of Washington, Box 351310, Seattle, Washington, 98195, USA.

five bursts were similar across many stations and had a move-out consistent with a single location (Supplementary Fig. 1). The relative timing of the tremor arrivals differed from that of the through-going Denali surface waves, so we can be sure that the tremor is not an artefact of filtering or an instrumental effect caused by large amplitudes. The simultaneous arrival of the Rayleigh waves and the tremor at station PHC is merely a coincidence, because their relative timing has not yet been corrected for different source locations and propagation velocities.

Our triggered tremor observations cover the broad range of frequencies 4–30 Hz with signal-to-noise ratio exceeding 10:1. In contrast, observations of tremor during an ETS episode ('ETS tremor') are strongest in a narrower band (1–10 Hz). We examined other reported incidences of earthquake-triggered tremor[16,17] and found similarly high signal amplitudes for a broad range of frequencies. While the visible bandwidth of triggered and ETS tremor is different, spectral comparison revealed that both have a similar high-frequency spectral shape (Supplementary Fig. 2), with the larger signal-to-noise ratio of the triggered tremor making it visible at a broader range of frequencies. It follows that ETS tremor and triggered tremor may be generated by the same physical process. We note that while the spectra of the triggered tremor diverges from the ETS tremor below 2 Hz, this difference arises because the triggered tremor is superposed on lower-frequency *S* coda and surface waves from the Denali earthquake. Like previous studies[2,3,9], we found that tremor is deficient in high frequencies relative to earthquakes.

We located the tremor by a search over a grid of possible locations, cross-correlating the envelopes from the four best recordings of a window containing the three largest bursts of tremor. This method is similar to that used by Obara[2]. We corrected the known-to-be-erroneous timing at stations HOLB and EDB by 13 s and 102 s respectively. These figures were determined from phase-arrival residuals of local earthquakes occurring on the same day as the Denali earthquake. The epicentral location of the tremor is tightly constrained to be 50.19° N ± 6 km and 127.61° W ± 4 km (Fig. 1). The source depth is loosely constrained to lie between 10 and 34 km. Epicentral locations obtained for the first and second bursts place them approximately 10 and 6 km to the southwest of the joint location for bursts 2 to 4. This suggests that the tremor may have migrated in a northeasterly direction. The triggered tremor is located further north and more up-dip than the majority of ETS tremor in the region[22]. This may indicate that slightly different conditions are needed to produce triggered tremor rather than ETS tremor.

To understand the triggering relationship between the surface waves and the tremor, we take the location of the tremor and shift the tremor and surface waves seismograms such that they directly reflect the timing relationship at the source location of the tremor (Fig. 3). The first four bursts of tremor were clearly triggered by the Love waves, but the Rayleigh waves arrived in time possibly to contribute to the fourth and fifth bursts. The tremor occurred when the Love wave displacement was southwest, but halted when the displacement reversed to the northeast. This is the expected polarity if the ambient stress field causes tremor, that is, when the Love waves from Denali amplified the shear stress in the direction of plate motion on the subduction interface. The amplitude of the triggered tremor also appears to scale with the up-dip shear stress of the Love waves resolved onto the plate interface.

The geometry of the Denali earthquake and Cascadia subduction zone is unique in that the surface waves from Denali propagated parallel to the strike of the subduction zone. This means that the Love waves are polarized such that their motions are parallel to the convergence direction and the Rayleigh wave motions oscillate parallel to the subduction zone strike. This geometry directs nearly all of the shear stress associated with the Love wave to alternately promoting and resisting plate convergence, while less of the stress associated with the Rayleigh wave directly affects plate convergence. This unique geometry and the fact that the Love waves are much larger than the Rayleigh waves explains why the Love waves triggered tremor and the Rayleigh waves did not.

We quantified the stresses and their timing on the subduction interface using synthetic seismograms that match the surface waves of the Denali earthquake on Vancouver Island (Supplementary Figs 3 and 4). In this calculation the Love waves caused the maximum shear stress resolved in the updip/downdip direction (43 kPa), while the maximum dilatational stress, which was caused by the Rayleigh waves, was only 12 kPa. The significantly larger shear stresses on the plate interface suggest that they were responsible for the triggering. For dilatation to be responsible for the tremor, it would require the tremor to be at least three times more sensitive to dilatation than shear. Even more compelling evidence for shear-triggered tremor is

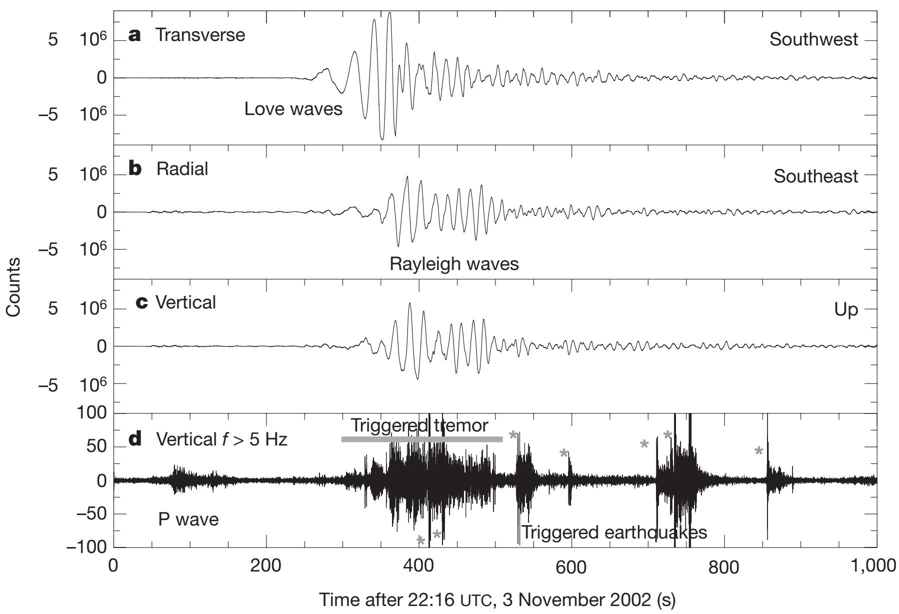


Figure 2 | Broadband, three-component recording of the Denali earthquake and locally triggered tremor and earthquakes. Record from station PHC showing the *P* wave and surface waves from Denali and locally triggered tremor and earthquakes. Raw records for transverse (**a**, 238°) radial (**b**, 148°) and vertical velocity (**c**). **d**, 5 Hz high-pass-filtered vertical velocity trace. Triggered local earthquakes are indicated by grey asterisks in the bottom panel.

that the timing of the tremor pulses and the shear stress increases match, while the Rayleigh waves arrive too late to trigger tremor.

The modest shear stress changes required to trigger tremor suggest several possible conditions within the subduction zone where the tremor occurred: the effective stress (the difference between the normal stress and pore pressure) is very low, the static friction and shear strength are low, and/or the stresses were near critical levels at the time of the Denali earthquake. Evidence from triggered earthquakes leads to similar suggestions. The existence of low effective stresses is consistent with earthquakes in geothermal fields (regions with hot fluids/gas that are presumably under high pressure) triggering under lower dynamic stresses (as low as 5 kPa; ref. 23) than in other crustal regions (of the order of hundreds to thousands of kilopascals; ref. 24). It is also consistent with a theoretical rate-state frictional model of ETS[25]. The recurrence interval of ~14 months between slow slip events in northern Cascadia and estimates showing they relax much (about two-thirds) of the stress accumulated during the interval[26] provide evidence for the second possibility: a weak interface.

We hypothesize that shear stress on the plate interface from the Love waves of the Denali earthquake triggered the tremor. Specifically, by analogy with ETS, we suggest that the shearing caused by the Love waves induced slow shear-failure on the plate interface, radiating the tremor we observe. Thus when the Love waves stressed the fault in the sense of the subduction deformation, the megathrust allowed an increment of subduction-directed fault slip, accompanied by tremor. In other words, tremor reflects a process in which failure follows a simple frictional behaviour. This model is consistent with the observations that the tremor is only triggered when the ambient stress field is amplified and that the tremor amplitude scales with the amplitude of the shear stresses associated with the triggering waves.

Our idea that triggered tremor follows a simple frictional threshold can also be used to explain the observations of triggered tremor that previous authors have attributed to dilatation and fluid flow[17,18]. For those cases where large dilatations are observed coincident with triggered tremor, we argue that the dilatations would also reduce normal stresses on the plate interface, thus increasing the likelihood of frictional failure.

Although we appeal to frictional failure as the physical mechanism responsible for generating the triggered tremor we observe, we still believe fluids to be important: we expect that high fluid pressures within the subduction zone do facilitate the slip. We suggest that the phenomenon we observed—low stress-drop or slow shear-failure in response to dynamically elevated shear stress—may also occur in locations closer to mainshock fault ruptures (where dynamic stresses are large), which implies that it may be important in the rupture process itself.

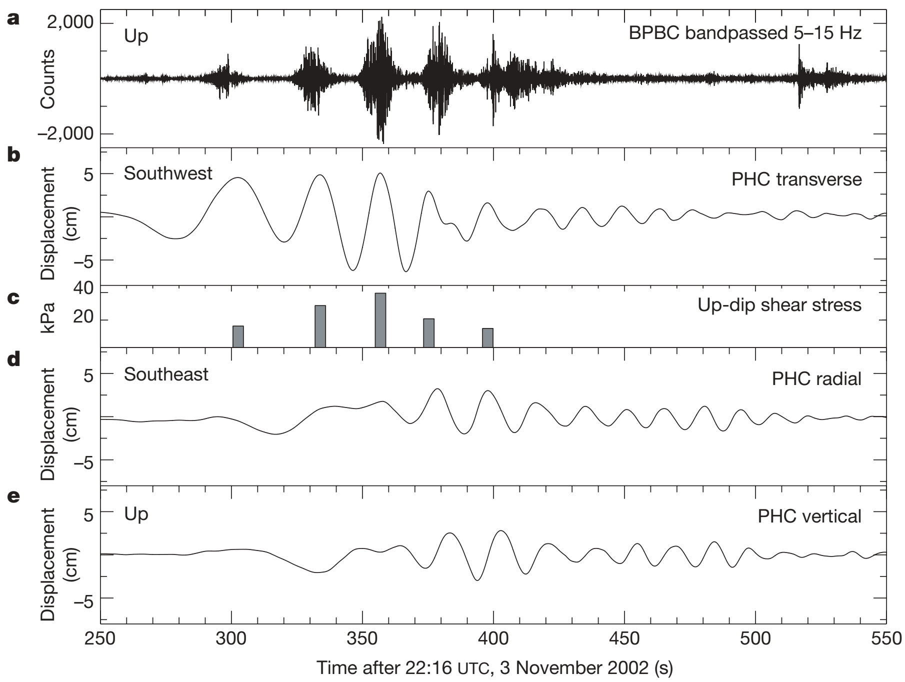


Figure 3 | Comparison of tremor and the surface waves that triggered it at the site where the tremor is radiated. **a**, Tremor at the closest station BPBC, time-adjusted by 5.14 s to reflect the travel time of *S* waves from the tremor source to BPBC. We used the epicentre determined for the window containing the second to fourth bursts of tremor and a depth of 15 km, the approximate depth of the plate interface, to estimate this travel time. Bursts one and two appear early because a single correction was applied to sources with different locations. Bursts 1 and 2 emanate from a source closer to BPBC than the sources corresponding to the bursts used to derive the correction; their locations would require shifts of ~1.0 and 0.8 s to the right, respectively, aligning them more closely with the triggering surface waves. **b**, **d**, **e**, Instrument-corrected displacement seismograms for the transverse, radial and vertical components at PHC, the closest three-component, broadband station to the tremor, shifted by 8.40 s seconds to reflect the difference in arrival time of the third and largest Love wave pulse at PHC and at the tremor source. The velocity for each triggering Love wave pulse was estimated using its relative arrival times at ten broadband stations on Vancouver Island and on the Canadian mainland. Because surface waves are dispersive—that is, their velocity depends on their wavelength—the correction we have applied is too large for the faster surface wave arrivals (for example, the first triggering pulse should be shifted left ~0.9 s) and too small for the slower surface wave arrivals (for example, the last triggering pulse should be shifted right ~0.5 s). The tremor appears to turn on when displacement is to the southwest and turn off when it is to the northeast, suggesting that the strains induced by the passage of the Love waves modulated the tremor. **c**, Predicted maximum shear stress on the plate interface from the five triggering Love wave pulses is plotted as a vertical bar centred on the maximum of each pulse. The shear stress values do not have a 1:1 relationship with the observed displacements because the stresses induced by a surface wave depend on the period of the waves. In this geometry, as the period of the Love waves increases, the shear stress on the plate interface decreases (Supplementary Fig. 5). Using a transfer function from surface displacement to shear stress on the plate interface determined from the synthetic seismograms (Supplementary Figs 3 to 5), we estimate the amount of shear stress on the plate interface by determining the period and maximum amplitude of the five triggering bursts from the displacement seismograms at PHC.

Received 28 February; accepted 13 June 2007.

1. McNutt, S. R. Volcanic seismology. *Annu. Rev. Earth Planet. Sci.* **33,** 461–491 (2005).
2. Obara, K. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* **296,** 1679–1681 (2002).
3. Kao, H. *et al.* A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* **436,** 841–844 (2005).
4. McCausland, W., Malone, S. & Johnson, D. Temporal and spatial occurrence of deep non-volcanic tremor: From Washington to northern California. *Geophys. Res. Lett.* **32,** L24311, doi:10.1029/2005GL024349 (2005).
5. Katsumata, A. & Kamaya, N. Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan. *Geophys. Res. Lett.* **30,** 1020, doi:10.1029/2002GL015981 (2003).
6. Seno, T. & Yamasaki, T. Low-frequency tremors, intraslab and interplate earthquakes in Southwest Japan—from a viewpoint of slab dehydration. *Geophys. Res. Lett.* **30,** 2171, doi:10.1029/2003GL018349 (2003).
7. Rogers, G. & Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* **300,** 1942–1943 (2003).
8. Obara, K., Hirose, H., Yamamizu, F. & Kasahara, K. Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophys. Res. Lett.* **31,** L26302, doi:10.1029/2004GL020848 (2004).
9. Shelly, D. R., Beroza, G. C. & Ide, S. Non-volcanic tremor and low frequency earthquake swarms. *Nature* **446,** 305–307 (2007).
10. Ide, S., Shelly, D. R. & Beroza, G. C. Mechanism of deep low frequency earthquakes: Further evidence that deep non-volcanic tremor is generated by shear slip on the plate interface. *Geophys. Res. Lett.* **34,** L03308, doi:10.1029/2006GL028890 (2007).
11. Shelly, D. R., Beroza, G. C., Ide, S. & Nakamula, S. Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* **442,** 188–191 (2006).
12. Ide, S., Beroza, G. C., Shelly, D. R. & Uchide, T. A scaling law for slow earthquakes. *Nature* **447,** 76–79 (2007).
13. Obara, K. & Hirose, H. Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* **417,** 33–51 (2006).
14. Royle, G. T., Calvert, A. J. & Kao, H. Observations of non-volcanic tremor during the northern Cascadia slow-slip event in February 2002. *Geophys. Res. Lett.* **33,** L18313, doi:10.1029/2006GL027316 (2006).
15. Obara, K. Time sequence of deep low-frequency tremors in the Southwest Japan Subduction Zone: Triggering phenomena and periodic activity. [in Japanese] *Chigaku Zasshi (J. Geogr.)* **112,** 837–849 (2003).
16. Miyazawa, M. & Mori, J. Detection of triggered deep low-frequency events from the 2003 Tokachi-oki earthquake. *Geophys. Res. Lett.* **32,** L10307, doi:10.1029/2005GL022539 (2005).
17. Miyazawa, M. & Mori, J. Evidence suggesting fluid flow beneath Japan due to periodic seismic triggering from the 2004 Sumatra-Andaman earthquake. *Geophys. Res. Lett.* **33,** L05303, doi:10.129/2005GL025087 (2006).
18. Miyazawa, M. & Brodsky, E. E. Fluid related deep low-frequency earthquakes resonant with the Rayleigh waves from the 2004 Sumatra-Andaman earthquake. *J. Geophys. Res.* (submitted).
19. Gomberg, J., Bodin, P., Larson, K. & Dragert, H. Earthquake nucleation by transient deformations caused by the $M=7.9$ Denali, Alaska, earthquake. *Nature* **427,** 621–624 (2004).
20. Tsuboi, S., Komatitsch, D., Ji, C. & Tromp, J. Broadband modeling of the 2002 Denali, Alaska Mw 7.9 earthquake on the Earth Simulator. *Phys. Earth Planet. Inter.* **139,** 305–313 (2003).
21. Mazzotti, S. *et al.* Current tectonics of northern Cascadia from a decade of GPS measurements. *J. Geophys. Res.* **108,** 2554, doi:10.1029/2003JB002653 (2003).
22. Kao, H., Shan, S.-J., Rogers, G. & Dragert, H. Migration characteristics of seismic tremors in the northern Cascadia margin. *Geophys. Res. Lett.* **34,** L03304, doi:10.1029/2006GL028430 (2007).
23. Brodsky, E. E. & Prejean, S. G. New constraints on mechanisms of remotely triggered seismicity at Long Valley Caldera. *J. Geophys. Res.* **110,** B04302, doi:10.1029/2004JB003211 (2005).
24. Gomberg, J., Reasenberg, P. A., Bodin, P. & Harris, R. A. Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* **411,** 462–466 (2001).
25. Liu, Y. & Rice, J. R. Spontaneous and triggered aseismic deformation transients in a subduction fault model. *J. Geophys. Res.* (submitted).
26. Dragert, H., Wang, K. & Rogers, G. Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth Planets Space* **56,** 1143–1150 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H. Kao, G. Rogers and H. Dragert provided comments and data that assisted with the tremor source location. Comments from W. McCausland, N. Beeler and H. Houston improved this manuscript. The data used in this study come from the Canadian National Seismograph Network and are distributed freely by The Geological Survey of Canada. This work was inspired by conversations with T. Pratt.

Author Contributions J.L.R. and J.E.V. found and identified the tremor, J.G. and P.B. computed the fault plane stresses, K.C.C. located the tremor source, and J.L.R. performed the spectral analysis. J.L.R., with the help of all, integrated the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.L.R. (justin@ess.washington.edu).

LETTERS

Jaws and teeth of the earliest bony fishes

Hector Botella[1], Henning Blom[2], Markus Dorka[3], Per Erik Ahlberg[2] & Philippe Janvier[4,5]

Extant jawed vertebrates, or gnathostomes, fall into two major monophyletic groups, namely chondrichthyans (cartilaginous fishes) and osteichthyans (bony fishes and tetrapods). Fossil representatives of the osteichthyan crown group are known from the latest Silurian period, 418 million years (Myr) ago, to the present. By contrast, stem chondrichthyans and stem osteichthyans are still largely unknown. Two extinct Palaeozoic groups, the acanthodians and placoderms, may fall into these stem groups or the common stem group of gnathostomes, but their relationships and monophyletic status are both debated. Here we report unambiguous evidence for osteichthyan characters in jaw bones referred to the late Silurian (423–416-Myr-old) fishes *Andreolepis hedei* and *Lophosteus superbus*, long known from isolated bone fragments, scales and teeth, and whose affinities to, or within, osteichthyans have been debated[1–11]. The bones are a characteristic osteichthyan maxillary and dentary, but the organization of the tooth-like denticles they bear differs from the large, conical teeth of crown-group osteichthyans, indicating that they can be assigned to the stem group. *Andreolepis* and *Lophosteus* are thus not only the oldest but also the most phylogenetically basal securely identified osteichthyans known so far.

Scales and minute bone fragments of *Lophosteus* and *Andreolepis* are classical fish remains from the Upper Silurian (423–416-Myr-old) limestone of the Baltic area[1–12]. They are now known elsewhere in the world, and *Lophosteus* occurs up to the early Devonian period (410 Myr ago)[8,9,13]. However, their affinities have long been a riddle. When first discovered, *Lophosteus* was tentatively referred to osteichthyans on the basis of its diamond-shaped scales[1,2], but its stellate ornamentation was later regarded as more suggestive of placoderms (armoured jawed vertebrates)[13] and its large fin spines suggested a relationship to acanthodians[5,8,10]. The latter are Ordovician to Permian (445–294-Myr-old) fishes that have been regarded as possible stem osteichthyans but lack the characteristic osteichthyan suite of large dermal bones and rows of shedding marginal teeth (acanthodian teeth, if present, are generally arranged in whorls or statodont rows; that is, non-shedding tooth rows growing by the addition of new teeth at one end). The relationships of *Andreolepis* have been more consensual. *Andreolepis* is generally regarded as a primitive actinopterygian (ray-finned fish)[3,6,9] on the basis of its ganoid scale structure (superimposed layers of dentine and enameloid), but it has also been supposed to share some derived characters with onychodontiform sarcopterygians (lobe-finned fishes) and acanthodians[7].

Although known from thousands of minute remains collected by dissolving limestone blocks in acid, *Lophosteus* and *Andreolepis* have until now yielded no skull-bone feature that would unambiguously anchor their relationships to the osteichthyan crown group. None of the dermal bones recorded so far unambiguously resembles a typical osteichthyan dermal bone, or shows the characteristic closed sensory-line canals of osteichthyans. No sensory-line bearing bone of *Andreolepis* is yet known, and those of *Lophosteus* only show broad sensory-line grooves[4,5,8,9]. Only one large dermal bone fragment of *Andreolepis* has been tentatively referred to a cleithrum, a dermal bone of the osteichthyan shoulder girdle[7]. We are therefore left with the ganoid scales of *Andreolepis*, including its possible fulcral scales that strengthened the leading edge of the fins, which are convincingly actinopterygian-like[6,7]. Nevertheless, no characteristic osteichthyan lepidotrich (bony dermal fin ray segment) has ever been recorded in association with either *Andreolepis* or *Lophosteus*, whereas they are common in Devonian (416–359-Myr-old) and younger osteichthyan-bearing sediments. This combination of characteristics indicates that both genera might belong to the osteichthyan stem group.

Two new discoveries of dermal jaw bones strongly support the placement of *Andreolepis* and *Lophosteus* in the osteichthyan stem group. Like all the previous material they are isolated fragments but can confidently be assigned to their respective genera on the basis of their dermal ornament (see Supplementary Information). First, among the incomplete skull bones of *Andreolepis* collected from the island of Gotland (Sweden; see Supplementary Information) some years ago, several have been identified as fragments of dentary; that is, the dermal bone of the lower jaw that normally bears the lateral series of large conical teeth ('outer dental arcade') in osteichthyans. We show here the best preserved fragment (Fig. 1a–d). This dentary curiously lacks large teeth, its edge bearing only rows of enlarged denticles that laterally grade into the rounded tubercles of the external bone ornamentation (Fig. 1d). Its internal surface is similar to that of the dentary in Devonian actinopterygians and sarcopterygians but, contrary to actinopterygian dentaries, there is no evidence for a mandibular sensory-line canal (Fig. 1a; see Supplementary Information). This suggests that a ventral row of infradentaries carried the canal, as in sarcopterygians.

Second, a large denticle-bearing dermal bone, ornamented with the characteristic stellate tubercles of *Lophosteus*, was discovered recently in late Silurian erratic boulders from Germany (see Supplementary Information). Its posteriorly increasing depth recalls an osteichthyan maxillary, and its denticles, although small, are slightly more distinct from the external ornamentation than in the dentary of *Andreolepis* (Fig. 2a–c). However, they differ from the characteristic maxillary teeth of early actinopterygians and sarcopterygians in being organized in at least three, almost parallel, longitudinal rows that, when combined, form transverse denticle files (Fig. 2d). Minute fragments of denticle-bearing dermal bones of *Lophosteus* (notably a probable 'denticle whorl') have been mentioned previously[4,5], but the specimen described here is clearly identifiable as a maxillary, as a result of the presence of a narrow medial horizontal lamina (Fig. 2b, d) and a low anterior prolongation that bears a small anterior overlap area (Fig. 2a). As a whole it compares in

[1]Departamento de Geología, Universitat de València, Dr Moliner 50, 46100 Burjassot, Valencia, Spain. [2]Evolutionary Organismal Biology, Department of Physiology and Developmental Biology, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden. [3]Humboldt-Universität zu Berlin, Museum für Naturkunde, Invalidenstrasse 43, D-10115 Berlin, Germany. [4]UMR 5143, CNRS, Département Histoire de la Terre, Muséum National d'Histoire Naturelle, 47 rue Cuvier, 75231 Paris Cedex 05, France. [5]Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

outline with the anterior part of the maxillary of early actinopterygians (Fig. 1e) and certain early sarcopterygians, notably onychodontiforms (Fig. 2e, f). The overlap area may be for a short lacrymal bone (Fig. 2a, f) and there is a possible suture area for the premaxillary (see Supplementary Information).

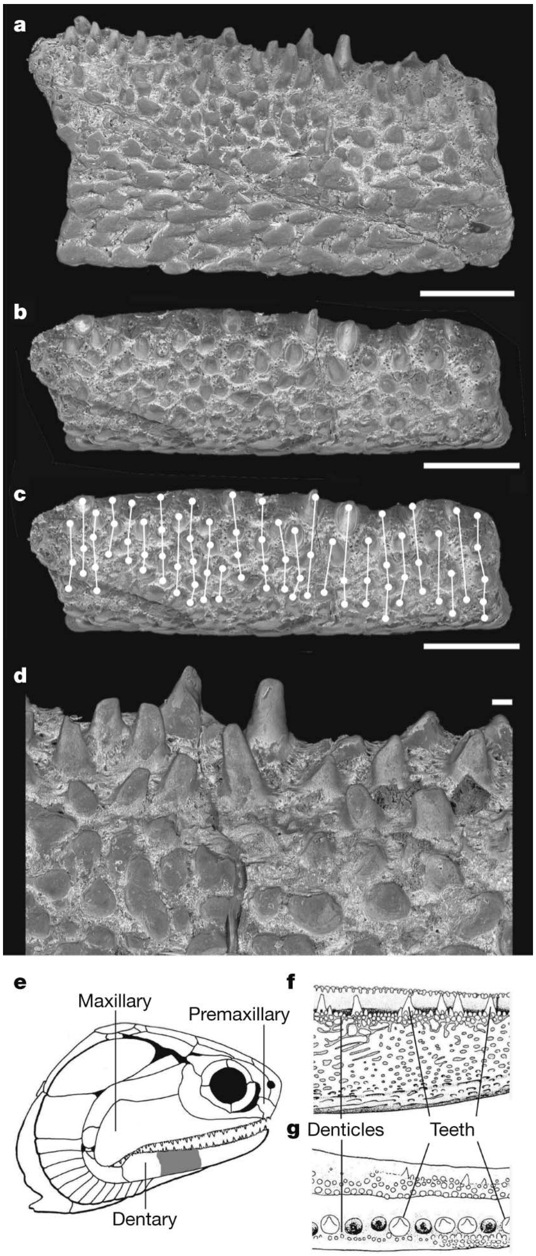


Figure 1 | Comparison of the dentary of *Andreolepis* with that of an actinopterygian. **a**–**c**, Incomplete right dentary of *Andreolepis hedei*, Upper Silurian (Ludlow) of Gannes, Gotland, Sweden (Lund University LOG87-301DF; scale bars, 1 mm (**a**–**c**) and 0.1 mm (**d**); scanning electron micrographs); **a**, Lateral view. **b**, **c**, Dorsal view showing the transition between the tooth-like denticles of the dorsal edge and the external ornamentation of the bone (**b**) and the possible arrangement of the denticles into files with alternating denticle initiation (**c**). **d**, Detailed view of the denticles in the middle part of the bone. **e**–**g**, Comparative illustration of the late Devonian (380-Myr-old) actinopterygian *Moythomasia* (from ref. 27; not to scale). **e**, Head showing the portion of the dentary (grey) corresponding approximately to the dentary fragment of *Andreolepis* in **a**–**d**. **f**, **g**, Lateral (**f**) and dorsal (**g**) views of the dentary showing the monolinear series of true conical teeth (absent in *Andreolepis*) and the labial series of denticles grading into the ornamentation, comparable to the denticles of *Andreolepis*.

The dentary of *Andreolepis* and the maxillary of *Lophosteus* both differ from their homologues in younger osteichthyans in lacking the series of large, conical teeth that are conspicuous in basal actinopterygians (Fig. 1f, g) and sarcopterygians (for example *Onychodus*; Fig. 2e, f), except for lungfishes, which have lost the marginal tooth rows[14]. In early actinopterygians and sarcopterygians, these large teeth are nevertheless bordered laterally by series of pointed denticles that gently grade into the external ornamentation (Fig. 1f, g) and are thus suggestive of the denticles on the dentary of *Andreolepis* and the maxillary of *Lophosteus*. In both forms, these denticles are similarly curved towards the jaw margin and show a marked gradient in size, the medialmost ones being larger than the lateral ones (Fig. 2d).

The presence of osteichthyan marginal jaw bones in *Andreolepis* and *Lophosteus*, coupled with the absence of the characteristic osteichthyan linear shedding tooth rows, is most readily interpreted as reflecting a phylogenetic position in the osteichthyan stem group (Fig. 3). Consequently, the ganoid scale structure of *Andreolepis*, long

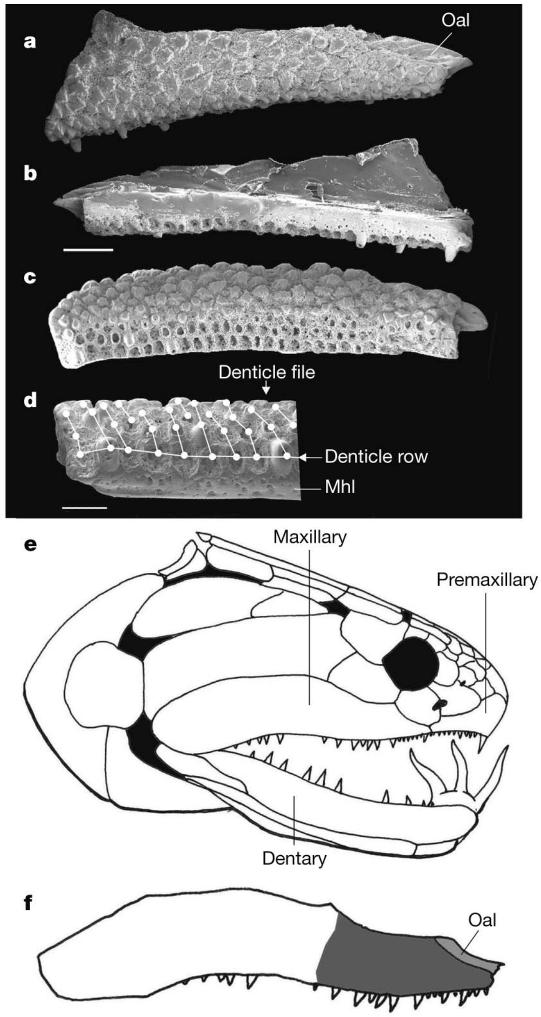


Figure 2 | Comparison of the maxillary of *Lophosteus* with that of a sarcopterygian. **a**–**d**, Right maxillary of *Lophosteus superbus*, Upper Silurian (Middle–Upper Pridoli) erratic boulder of Germany (Museum für Naturkunde, Berlin MB.f.17035; scale bars, 1 mm (**a**–**c**) and 0.1 mm (**d**); scanning electron micrographs). **a**, Lateral view. **b**, Medial view. **c**, Ventral view showing the transition between the denticles and the external ornamentation. **d**, Detail of the posterior end of the ventral edge showing the organization of the denticles into rows and files. **e**, **f**, Comparative illustration of the late Devonian (380-Myr-old) onychodontiform sarcopterygian *Onychodus* (from ref. 28, not to scale). **e**, Head in lateral view. **f**, Maxillary showing the portion (grey) approximately corresponding to the maxillary fragment of *Lophosteus* in **a**–**d**. Mhl, medial horizontal lamina; OaL, overlap area for the lacrymal.

regarded as an actinopterygian 'signature', is probably a more general character than previously believed, as also suggested by its presence in the supposed early Devonian actinopterygian *Dialipina*, whose anatomy is in fact strongly at odds with that of all early crown osteichthyans[15]. Moreover, the skull bones of the basalmost sarcopterygian *Meemannia* also possess a ganoine-like covering[16]. *Lophosteus*, *Andreolepis* and, potentially, *Dialipina* thus begin to define a character complement for the upper part of the osteichthyan stem group[17]. This will form a valuable basis for evaluating the relationships of forms such as acanthodians whose suggested similarities to osteichthyans are more tenuous. *Lophosteus* and *Andreolepis* also illuminate the evolution of the osteichthyan dentition.

Several theories have been proposed for the origin of gnathostome tooth patterning and replacement[18–23]. A recently favoured hypothesis suggests that the initial tooth patterning and replacement mechanisms are modular and derived from the pattern of the pharyngeal denticle whorls found in some fossil jawless vertebrates[19,20]. This accounts for the tooth replacement of sharks and early chimaeroids, and for the parasymphysial tooth whorls of acanthodians and primitive osteichthyans. In the tooth rows of the most primitive living actinopterygian, the bichir (*Polypterus*), and in living amphibians, new teeth grow medially to older ones in a transverse sequence. This mechanism is interpreted as primitive for osteichthyans[18], suggesting that patterning and replacement of these tooth rows are also fundamentally modular[21–23]. The organization of the jaw denticles of *Lophosteus* into parallel longitudinal rows and transverse files, which recalls that of the tooth 'families' (or files) of chondrichthyans and certain acanthodians[24], may thus illustrate an early stage of osteichthyan marginal tooth patterning. It is possible that each of these denticle files corresponds to an individual tooth 'capsule'; that is, a functional module[22,23]. The largest (and presumably youngest) denticles form the medialmost row, and new, larger and larger denticles appeared at the medial end of each denticle file as the maxillary grew. Although the denticles seem to grade into the tubercles of the external bone ornamentation, their pattern differs from the random distribution of the latter (Fig. 2c). The denticles of *Andreolepis* show the same medial-to-lateral size gradient as in *Lophosteus* and seem to be arranged in transverse files with alternating initiation of new denticles (Fig. 1b, c), suggesting a tooth initiation pattern somewhat like that in sharks[19].

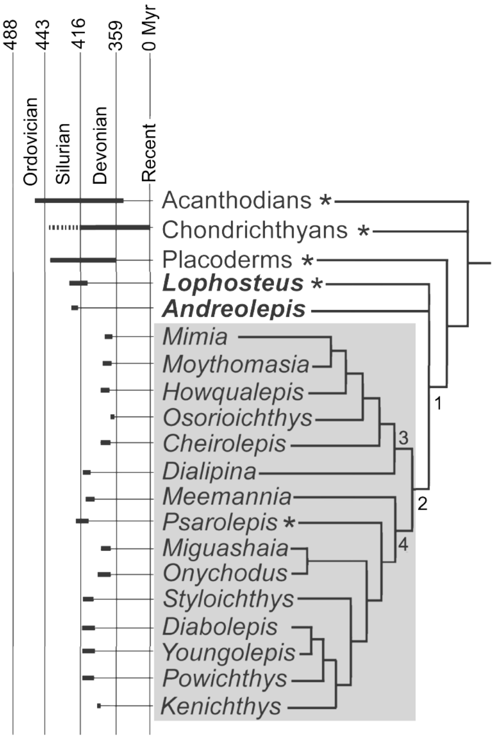


Figure 3 | Relationships of *Andreolepis* and *Lophosteus* to other jawed vertebrates. The diagram is taken from the tree in ref. 16, with *Andreolepis* and *Lophosteus* added; bold lines show distribution through time. The debated chondrichthyan record is dashed. Crown-group osteichthyans (grey) include here only the Devonian relatives of major living taxa, for example actinopterygians (*Mimia*/*Moythomasia*), coelacanths (*Miguashaia*), dipnoans (*Diabolepis*) and tetrapodomorphs (*Kenichthys*). Asterisks indicate taxa with fin spines or spinal plates. 1–4, Jawbone characters: 1, osteichthyans (denticles on edge of dentary, maxillary and possibly premaxillary); 2, crown-group osteichthyans (broad medial horizontal lamina of dentary and maxillary bearing larger monolinear tooth row); 3, actinopterygians (sensory-line canal in dentary); 4, sarcopterygians (largest teeth ('fangs') on coronoids, ectopterygoids, dermopalatine and vomers).

The dentition of *Lophosteus* is more obviously patterned in rows and files than that of *Andreolepis* and is therefore closer to the condition in crown osteichthyans: it could in principle be converted into a typical osteichthyan tooth row by introducing a shedding phase such that each tooth in a file was shed before its more mesial replacement was implanted. In contrast, the diversity and size of the fin spines referred to *Lophosteus* suggests a more primitive condition than *Andreolepis*, recalling that of acanthodians[4,5,8,10]. Paired and unpaired fin spines are also known in the early Devonian sarcopterygian *Psarolepis*[25], as well as in the earliest chondrichthyans[26], and are probably a general character of the crown-group gnathostomes, or even all jawed vertebrates if regarded as homologous to the spinal plates of placoderms (Fig. 3). Which of *Andreolepis* or *Lophosteus* is more closely related to crown-group osteichthyans cannot be determined until more complete specimens are found. Recent discoveries of the earliest crown-group osteichthyans from the early Devonian of China have considerably increased our understanding of the actinopterygian–sarcopterygian divergence and overturned some supposedly diagnostic characters of these groups[16]. Even though *Andreolepis* and *Lophosteus* possessed characteristic osteichthyan jaw bones with incipient osteichthyan tooth organization, the rest of their skull pattern may turn out to be quite different from that of crown-group osteichthyans.

Received 7 March; accepted 5 June 2007.

1. Pander, C. H. *Monographie der fossilen Fische des silurischen Systems der Russisch-Baltischen Gouvernements* (Mem. Akad. Wiss., St Petersburg, 1856).
2. Rohon, J. V. Die Obersilurischen Fische von Oesel. Theil 2. *Mém. Acad. Imp. Sci. St Petersbourg* **41**, 1–124 (1893).
3. Gross, W. Fragliche Actinopterygier-Schuppen aus dem Silur Gotlands. *Lethaia* **1**, 184–218 (1968).
4. Gross, W. *Lophosteus superbus* Pander, ein Teleostome aus dem Silur Oesels. *Lethaia* **2**, 15–47 (1969).
5. Gross, W. *Lophosteus superbus* Pander: Zähne, Zahnknochen und besondere Schuppenformen. *Lethaia* **4**, 131–152 (1971).
6. Schultze, H.-P. Ausgangsform und Entwicklung der rhombischen Schuppen der Osteichthyes (Pisces). *Paläont. Z.* **51**, 152–168 (1977).
7. Janvier, P. On the oldest known teleostome fish *Andreolepis hedei* Gross (Ludlow of Gotland), and the systematic position of the lophosteids. *Eesti NSV Teaduste Akadeemia Toimetised. Geologia* **3**, 88–95 (1978).
8. Schultze, H.-P. & Märss, T. Revisiting *Lophosteus*, a primitive osteichthyan. *Acta Univ. Latv.* **674**, 57–78 (2004).
9. Märss, T. *Andreolepis* (Actinopterygii) in the Upper Silurian of northern Eurasia. *Proc. Estonian Acad. Sci. Geol.* **50**, 174–189 (2001).
10. Otto, M. Zur systematischen Stellung der Lophosteiden (Obersilur, Pisces inc. sedis). *Paläont. Z.* **65**, 345–350 (1991).
11. Märss, T. Silurian vertebrates of Estonia and West Latvia. *Fossilia Baltica* **1**, 1–104 (1986).
12. Fredholm, D. Vertebrates in the Ludlovian Hemse Beds of Gotland, Sweden. *Geol. Fören. Stockh. Förhandl.* **110**, 157–179 (1988).
13. Burrow, C. A new lophosteiform (Osteichthyes) from the Lower Devonian of Australia. *Geobios Mem. Spec.* **19**, 327–333 (1995).
14. Ahlberg, P. E., Smith, M. M. & Johanson, Z. Developmental plasticity and disparity in early dipnoans (lungfish). *Evol. Dev.* **8**, 331–349 (2006).
15. Schultze, H.-P. & Cumbaa, S. L. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 315–332 (Taylor & Francis, London, 2001).

16. Zhu, M., Yu, X., Wang, W., Zhao, W. & Jia, L. A primitive fish provides key characters bearing on deep osteichthyan phylogeny. *Nature* **44,** 77–80 (2006).
17. Friedman, M. & Blom, H. A new actinopterygian from the Famennian of East Greenland and the interrelationships of Devonian ray-finned fishes. *J. Paleontol.* **80,** 1186–1204 (2006).
18. Clemen, G., Bartsch, P. & Wacker, K. Dentition and dentigerous bones in juveniles and adults of *Polypterus senegalus* (Cladistia, Actinopterygii). *Ann. Anat.* **189,** 211–221 (1998).
19. Smith, M. M. & Coates, M. I. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 133–151 (Taylor & Francis, London, 2001).
20. Smith, M. M. Vertebrate dentitions and the origin of jaws: when and how pattern evolves. *Evol. Dev.* **5,** 394–413 (2003).
21. Huysseune, A. & Witten, P. E. Developmental mechanisms underlying tooth patterning in continuously replacing osteichthyan dentition. *J. Exp. Zool. (Mol. Dev. Evol.)* **306B,** 204–215 (2006).
22. Fraser, G. J., Graham, A. & Smith, M. M. Developmental and evolutionary origins of the vertebrate dentition: molecular controls for spatio-temporam organisation of tooth sites in osteichthyans. *J. Exp. Zool. (Mol. Dev. Evol.)* **306B,** 183–203 (2006).
23. Fraser, G. J., Berkovitz, B. K., Graham, A. & Smith, M. M. Gene deployment for tooth replacement in the rainbow trout (*Oncorhynchus mykiss*): a developmental model for evolution of the osteichthyan dentition. *Evol. Dev.* **8,** 446–457 (2006).
24. Schultze, H. P. & Zidek, J. Ein primitiver Acanthodier (Pisces) aus dem Unterdevon Lettlands. *Paläont. Z.* **56,** 95–105 (1982).
25. Zhu, M., Yu, X. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397,** 607–610 (1999).
26. Miller, R. F., Cloutier, R. & Turner, S. The oldest articulated chondrichthyan from the Early Devonian period. *Nature* **425,** 501–504 (2003).
27. Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia*, from the Upper Devonian of Western Australia. *Bull. Br. Mus. Nat. Hist.* **37,** 173–428 (1984).
28. Andrews, M., Long, J., Ahlberg, P., Barwick, R. & Campbell, K. The structure of the sarcopterygian *Onychodus jandemarrai* n. sp. from Gogo, Western Australia: with a functional interpretation of the skeleton. *Trans. R. Soc. Edinb. Earth Sci.* **96,** 197–307 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Lund University and the Institut für Geowissenschaften Tübingen for lending us specimens.

Author Contributions M.D. discovered the maxillary of *Lophosteus*, H. Blom restored the dentary of *Andreolepis*, and both provided stratigraphical data. All authors contributed equally to the rest of this study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.J. (janvier@mnhn.fr).

LETTERS

Morphological evolution through multiple *cis*-regulatory mutations at a single gene

Alistair P. McGregor[1]†, Virginie Orgogozo[1]†, Isabelle Delon[2]†, Jennifer Zanet[2], Dayalan G. Srinivasan[1], François Payre[2] & David L. Stern[1]

One central, and yet unsolved, question in evolutionary biology is the relationship between the genetic variants segregating within species and the causes of morphological differences between species. The classic neo-darwinian view postulates that species differences result from the accumulation of small-effect changes at multiple loci. However, many examples support the possible role of larger abrupt changes in the expression of developmental genes in morphological evolution[1–3]. Although this evidence might be considered a challenge to a neo-darwinian micromutationist view of evolution, there are currently few examples of the actual genes causing morphological differences between species[4–10]. Here we examine the genetic basis of a trichome pattern difference between *Drosophila* species, previously shown to result from the evolution of a single gene, *shavenbaby* (*svb*), probably through *cis*-regulatory changes[6]. We first identified three distinct *svb* enhancers from *D. melanogaster* driving reporter gene expression in partly overlapping patterns that together recapitulate endogenous *svb* expression. All three homologous enhancers from *D. sechellia* drive expression in modified patterns, in a direction consistent with the evolved *svb* expression pattern. To test the influence of these enhancers on the actual phenotypic difference, we conducted interspecific genetic mapping at a resolution sufficient to recover multiple intragenic recombinants. This functional analysis revealed that independent genetic regions upstream of *svb* that overlap the three identified enhancers are collectively required to generate the *D. sechellia* trichome pattern. Our results demonstrate that the accumulation of multiple small-effect changes at a single locus underlies the evolution of a morphological difference between species. These data support the view that alleles of large effect that distinguish species may sometimes reflect the accumulation of multiple mutations of small effect at select genes.

Differences in larval trichome pattern between *Drosophila* species offer an attractive model of morphological evolution. Over the past 30 years, numerous studies have identified upstream patterning[11,12] and downstream effector[13] genes regulating trichome development in *D. melanogaster*. Questions about the evolution of trichome patterns can therefore be formulated explicitly within a developmental framework.

Although the pattern of ventral trichomes has been conserved for more than 60 Myr, new dorsal trichome patterns have evolved repeatedly[6,14,15]. In most species of the *D. melanogaster* subgroup, the dorsal and lateral surface displays stout trichomes on 1° and 3° cells and naked 2° cells, and a lawn of fine trichomes on 4° cells[6] (Fig. 1). *D. sechellia* has evolved a trichome pattern in which 4° trichomes were replaced by naked cuticle (Fig. 1h, l)[6]. Interspecific whole-genome genetic mapping demonstrated that the *D. sechellia* 'naked' phenotype is recessive to the 'hairy' phenotype of other species and mapped this evolutionary change to a single X-linked gene, *shavenbaby/ovo* (*svb*)[6]. *Svb* is required cell-autonomously for trichome formation[12] and encodes a transcription factor regulating several classes of effector genes, which collectively build trichomes[13].

In *D. melanogaster*, *D. simulans* and *D. mauritiana*, *svb* is expressed in 1° and 3° dorsal cells and 4° dorsal and lateral cells[16] (Fig. 1). In *D. sechellia*, *svb* is expressed in 1° and 3° dorsal cells but not in the 4° cells (Fig. 1b, d, f), where trichomes are absent[6]. Together with previous genetic analyses, these expression patterns suggest that changes in the *cis*-regulatory region of *svb* underlie this evolved morphological pattern[6].

We therefore sought to identify enhancers that drive *svb* expression. We made a systematic series of *D. melanogaster* reporter constructs, from 50 kilobases (kb) upstream to 20 kb downstream of the first exon of *svb* (Fig. 2a, Supplementary Fig. 1 and Supplementary Table 1). We precisely mapped reporter expression by double-staining for Miniature, the product of a cell-autonomous target of *svb*[13] that accumulates in trichomes (Fig. 2b–g and Supplementary Fig. 2).

We found that three genomic regions drove expression in the epidermis, just before trichome differentiation. Each element contributes to both evolutionarily conserved and evolutionarily derived expression patterns. Dorsal expression of the 'proximal' enhancer started in stage 13 embryos, in 1° and 3° cells (Fig. 2d, and Supplementary Figs 3c, f and 4c, g, k). Beginning at stage 15, expression was observed in some dorsal, but not dorsolateral, 4° cells (Supplementary Figs 3f and 4g). The 'medial' enhancer drove expression in the dorsal 4° cells at stage 13 and later expanded into dorsolateral 4° cells (Fig. 2c, and Supplementary Figs 3b, e and 4b, f, j). The 'distal' enhancer drove expression in thoracic dorsal stripes and lateral 4° cells, starting at stage 14 and strengthening later (Fig. 2b, and Supplementary Figs 3a, d and 4a, e, i). In ventral trichome-producing cells, the proximal and medial enhancers drove strong expression and the distal enhancer drove weak expression (Supplementary Figs 3a–f and 5a–k). The epidermal expression of *svb* therefore seems to be regulated in a complex manner by three separable *cis*-regulatory elements spread over 50 kb.

To determine whether these enhancers have evolved in *D. sechellia*, we identified orthologous *D. sechellia* regions (which differ by 3–5% from the *D. melanogaster* sequences) and assayed their activity as transgenes in *D. melanogaster* (Supplementary Table 1). The *D. sechellia* 'proximal' enhancer drove expression in 1° and 3° dorsal cells in a pattern similar to that of the *D. melanogaster* 'proximal' enhancer (Fig. 2g, and Supplementary Figs 3i, l and 4n, r, v). However, unlike the *D. melanogaster* enhancer, expression from the

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. [2]Centre de Biologie du Développement, Bâtiment 4R3, 118 Route de Narbonne, 31062 cedex 4 Toulouse, France. †Present addresses: Institute for Genetics, University of Cologne, D-50674, Cologne, Germany (A.P.McG.); Université Pierre et Marie Curie, Bâtiment A, 7 Quai Saint Bernard, 75005 Paris, France (V.O.); The Gurdon Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 1QN, UK (I.D.).

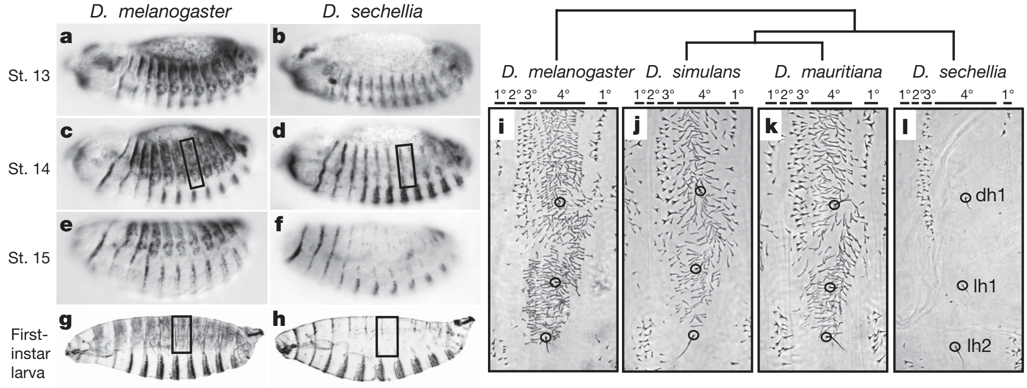


Figure 1 | Modified *svb* expression underlies the evolved trichome pattern of *D. sechellia*. a–f, *In situ* hybridization of *svb* transcript in stage (st.) 13, 14 and 15 *D. melanogaster* (**a**, **c**, **e**) and *D. sechellia* (**b**, **d**, **f**) embryos (dorsal up; anterior to the left). **g**, **h**, Trichome patterns on first-instar larvae of *D. melanogaster* (**g**) and *D. sechellia* (**h**). Black rectangles identify similar locations on embryos (**c**, **d**) and cuticles (**g**, **h**). **i–l**, Cuticle pattern of *D. melanogaster* (**i**), *D. simulans* (**j**), *D. mauritiana* (**k**) and *D. sechellia* (**l**). A phylogeny of the four species[27] is shown above; the clade is 2–3 Myr old[28]. All closely related outgroup species have a *D. melanogaster*-like trichome pattern (D.L.S., unpublished observations), indicating that the *D. sechellia* trichome pattern is derived. Positions of 1°, 2°, 3° and 4° cell types[11] are indicated, and the dorsal hair 1 (dh1), lateral hair 1 (lh1) and lateral hair 2 (lh2) are circled in black.

D. sechellia proximal enhancer was never observed in dorsal 4° cells. Expression of the *D. sechellia* 'medial' enhancer was restricted to dorsal 4° cells (Fig. 2f, and Supplementary Figs 3h, k and 4m, q, u). In contrast with the *D. melanogaster* medial enhancer, expression of the *D. sechellia* enhancer started later and did not extend to the lateral region. The *D. sechellia* 'distal' enhancer drove expression in thoracic stripes, in a similar manner to the *D. melanogaster* enhancer, but expression was observed only in restricted lateral spots (Fig. 2e, and Supplementary Figs 3g, j and 4l, p, t). At the time of trichome formation, each *D. sechellia* enhancer drove a ventral expression pattern similar to that of its *D. melanogaster* counterpart (compare Supplementary Fig. 5i–k with Supplementary Fig. 5t–v).

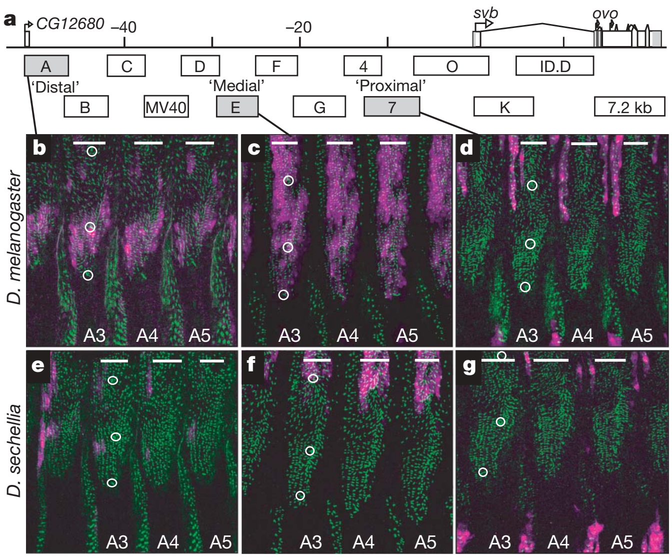


Figure 2 | Three enhancer regions, which collectively recapitulate the *svb* expression pattern, have evolved in *D. sechellia*. a, Genomic organization of the *ovo*/*svb* gene and regulatory regions up to *CG12680*, a gene located 50 kb upstream of the first *svb* exon. A systematic series of enhancer constructs, illustrated below the map, was generated (Supplementary Table 1) and only regions labelled as distal (A), medial (E) and proximal (7) drove epidermal expression at the time of trichome formation. **b–g**, Reporter patterns of enhancer constructs (purple) from *D. melanogaster* (**b–d**) and *D. sechellia* (**e–g**) were revealed by double staining of stage-15 embryos with anti-Min to stain trichomes (green) (Supplementary Fig. 2). Images show a lateral view of A3–A5 segments (bars); white circles indicate dh1, lh1 and lh2 sensory bristles. The *Dm*_distal construct (**b**) drives expression in lateral patches of 4° trichomes, *Dm*_medial (**c**) is strongly expressed in dorsal and dorsolateral 4° cells, and *Dm*_proximal (**d**) first drives expression in dorsal 1° and 3° cells and later in the dorsal-most 4° cells (Supplementary Figs 3 and 4). Although still expressed in the thorax, the *Ds*_distal enhancer (**e**) is strongly reduced in lateral 4° cells in the abdomen. At the time of trichome formation, the *Ds*_medial enhancer (**f**) is restricted to the dorsal-most 4° cells. The *Ds*_proximal enhancer (**g**) drives expression in dorsal 1° and 3° cells but was never observed at later stages in 4° cells (Supplementary Figs 3 and 4).

These results show that all three *svb* enhancers have evolved in *D. sechellia* and that these changes reflect a precise loss of expression in 4° cells. In addition, the *D. sechellia* medial and distal enhancers retain some activity in 4° cells, indicating that sites outside these regions might be required to repress this activity. Finally, minor changes were observed in the conserved 1° and 3° dorsal cells, and in ventral cells (Supplementary Fig. 5). These results suggest that evolution of the *D. sechellia svb* expression pattern was caused by multiple changes of limited effect rather than by drastic elimination of entire enhancers.

To test the actual function of these enhancers within their native genomic locations for patterning trichomes, we performed high-resolution interspecific recombination mapping[17]. We designed a two-step screen to maximize the probability of identifying recombinants within the *svb* gene. We first screened for recombinants between visible markers that flanked *svb* by about 1.2×10^6 base pairs (bp)[18] and then scored these selected individuals with molecular markers to identify 50 individuals with recombination breakpoints within the *svb* locus (Fig. 3, and Supplementary Information). This experiment provided a resolution of about one recombination breakpoint every 2 kb.

Recombinants that included the entire region upstream of the first *svb* exon from *D. mauritiana* produced trichome patterns indistinguishable from those of *D. mauritiana* (Fig. 3a). Conversely, chromosomes with the upstream *svb* region from *D. sechellia* produced a *D. sechellia*-like trichome pattern (Fig. 3e). These results demonstrate that the change(s) responsible for evolution of the *D. sechellia* phenotype are restricted to the genomic region that contains the three identified enhancers.

If the *D. sechellia* trichome pattern resulted from the evolution of a single site, then only *sechellia*-like and *mauritiana*-like phenotypes would have been observed. Instead, we identified three additional phenotypic classes. First, recombinants that included only the proximal enhancer from *D. mauritiana* produced a few dorsal 4° trichomes (intermediate type 1, Fig. 3b). Second, a chromosome including the medial and proximal enhancers from *D. mauritiana* produced a dense pattern of 4° trichomes in the dorsal and dorsolateral region (intermediate type 2, Fig. 3c). Last, chromosomes that

included only the distal enhancer from *D. mauritiana* produced a patch of dorsolateral and a few dorsal 4° trichomes (intermediate type 3; Fig. 3d). Backcrossing of all viable recombinant lines further ruled out any detectable influence of genomic regions outside *svb* on trichome patterns (Supplementary Table 3).

These genetic results prove that at least three separate changes have evolved in the *svb* upstream region to cause trichome loss in *D. sechellia*. Furthermore, the recombination breakpoints localize functionally evolved sites to genomic positions containing enhancers defined by reporter constructs. The distal *svb* enhancer element includes *CG12680*, which has the potential to encode a short peptide. However, this gene is unlikely to contribute to the evolved difference because *CG12680* is not expressed in embryos (data not shown) and complementation assays implicate *svb* alone as the causal determinant[6]. Finally, the recombinant intermediate phenotypes are similar to the expression patterns of the three individual enhancers (Supplementary Fig. 6). Our combined results therefore imply that each enhancer contains at least one genetic change. These changes may have occurred sequentially by loss of expression from the distal, medial and proximal enhancers, or in any other order (Fig. 4).

Given that laboratory-induced mutations in dozens of genes alter trichome patterns[11,12,19–22], it is striking that multiple mutations at a single locus account for the entire evolved difference. *Svb* seems peculiar in the network of genetic interactions that establish the trichome pattern, because it sits at the nexus of the upstream patterning genes and the downstream effector genes[13,23]. Although trichome pattern could be changed by altering any of several upstream genes, these changes would probably produce pleiotropic effects on other developmental processes. In contrast, none of the known downstream genes is sufficient on its own to prevent or promote trichome formation. Thus, changes at *svb* enhancers may provide the only available genetic mechanism to evolve trichome patterns without pleiotropic consequences.

Our results provide experimental evidence that the conflicting views of micromutationism and macromutationism can actually reflect observations of the same molecular mechanisms at different levels of resolution. Specifically, genes at integrative positions in developmental networks may be genetic 'hotspots' for evolutionary changes that differentiate species, although the individual mutations contributing to this change may be of smaller effect. Although results recently obtained from a broad range of species are consistent with this interpretation[4,24–26], only additional fine-scale functional analyses of morphological differences between species will allow a robust test of this hypothesis.

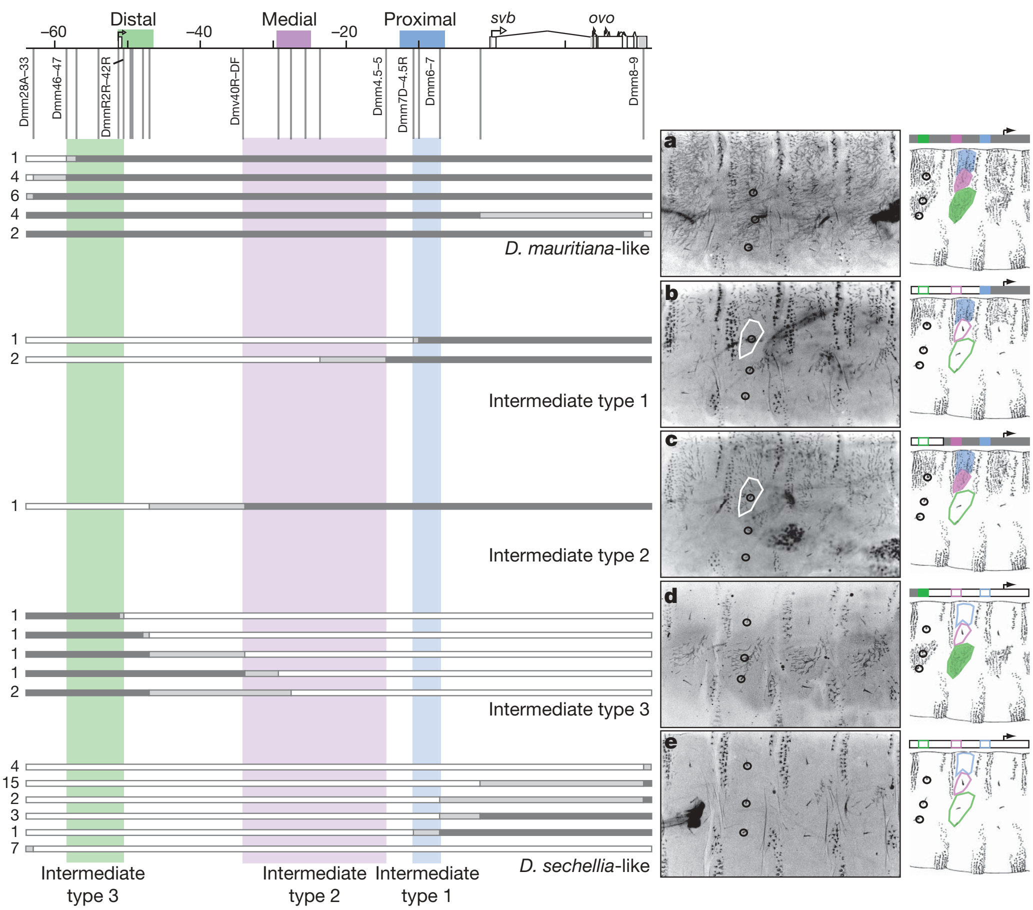


Figure 3 | High-resolution interspecific recombination mapping identifies three enhancer regions of *svb* that caused evolution of the *D. sechellia* trichome pattern. Vertical lines indicate the positions of molecular markers (Supplementary Tables 2 and 3) used to map recombinant chromosomes, grouped according to their trichome pattern. The number of flies carrying a particular recombination breakpoint are indicated on the left. The identity of source DNA in each recombinant chromosome is indicated by the colour of the bar: white, *D. sechellia* DNA; light grey, breakpoint region; darker grey, *D. mauritiana* DNA. **a**, Recombinants that included *D. mauritiana* DNA encompassing the three enhancers produced a *D. mauritiana*-like pattern. **b**, Recombinants containing only the *Dm*_proximal enhancer produced an intermediate type 1 phenotype, characterized by sparse dorsal 4° trichomes. **c**, A recombinant containing the *Dm*_proximal and *Dm*_medial enhancers produced an intermediate type 2 phenotype with dense dorsal and dorsolateral (within white outline) 4° trichomes. **d**, Recombinants containing only the *Dm*_distal enhancer produced intermediate type 3 patterns with lateral patches of 4° trichomes. **e**, Recombinants with *D. sechellia* DNA throughout the *svb cis*-regulatory region produced a *D. sechellia*-like trichome pattern. Diagrams of the trichome patterns are shown at the right, with *D. mauritiana* and *D. sechellia* enhancers indicated as filled and open rectangles, respectively.

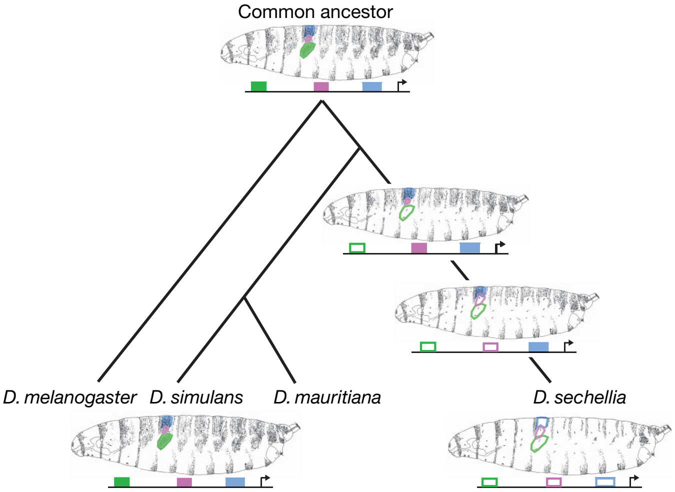


Figure 4 | A possible model of the evolutionary path of the *D. sechellia* trichome pattern. The three *svb* enhancers that drive *D. melanogaster*-like expression in the presumptive common ancestor (top) are indicated as filled rectangles. Each enhancer also drives expression in conserved dorsal and ventral domains, but only the evolving expression patterns are illustrated. At least one change per enhancer region has occurred independently to eliminate *svb* expression progressively in the dorsal and lateral 4° cells as illustrated, or in any other temporal sequence.

METHODS SUMMARY

Reporter constructs. Fragments of *svb* from *D. melanogaster* and *D. sechellia* (Fig. 2 and Supplementary Table 1) were cloned into standard reporter-gene constructs. Expression driven by reporter transgenes was assayed by *in situ* hybridization and by double stainings with an antibody recognizing the reporter gene product and antibodies against endogenous *Drosophila* proteins.

Recombination mapping. *D. mauritiana* flies carrying visible markers flanking the *svb* gene were crossed to *D. sechellia* flies. About 16,000 backcross progeny were examined for the presence of only one of the two flanking markers. About 600 recombinant female progeny were selected, and their wings were removed for genotyping with molecular markers flanking the *svb* locus (Supplementary Table 2). Females with a recombination event in the *svb* gene were then crossed to *D. sechellia* males and the resulting progeny were scored for their larval cuticle phenotypes. Recombination breakpoints within the *svb* gene were mapped by using additional molecular markers (Supplementary Table 2).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 November 2006; accepted 5 June 2007.
Published online 15 July 2007.

1. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Malden, 2001).
2. Marcellini, S. & Simpson, P. Two or four bristles: functional evolution of an enhancer of *scute* in Drosophilidae. *PLoS Biol.* **4,** e386 (2006).
3. Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305,** 1462–1465 (2004).
4. Wang, X. & Chamberlin, H. M. Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48 ovo* gene. *Genes Dev.* **16,** 2345–2349 (2002).
5. Doebley, J., Stec, A. & Gustus, C. *teosinte branched1* and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics* **141,** 333–346 (1995).
6. Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo/shavenbaby*. *Proc. Natl Acad. Sci. USA* **97,** 4530–4534 (2000).
7. Stern, D. L. A role of *Ultrabithorax* in morphological differences between *Drosophila* species. *Nature* **396,** 463–466 (1998).
8. Yoon, H. S. & Baum, D. A. Transgenic study of parallelism in plant morphological evolution. *Proc. Natl Acad. Sci. USA* **101,** 6524–6529 (2004).
9. Wang, H. *et al.* The origin of the naked grains of maize. *Nature* **436,** 714–719 (2005).
10. Hay, A. & Tsiantis, M. The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nature Genet.* **38,** 942–947 (2006).
11. Bokor, P. & DiNardo, S. The roles of *hedgehog*, *wingless* and *lines* in patterning the dorsal epidermis in *Drosophila*. *Development* **122,** 1083–1092 (1996).
12. Payre, F., Vincent, A. & Carreno, S. *ovo/svb* integrates Wingless and DER pathways to control epidermis differentiation. *Nature* **400,** 271–275 (1999).
13. Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. & Plaza, S. Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol.* **4,** e290 (2006).
14. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424,** 935–938 (2003).
15. Khila, A., El Haidani, A., Vincent, A., Payre, F. & Souda, S. I. The dual function of *ovo/shavenbaby* in germline and epidermis differentiation is conserved between *Drosophila melanogaster* and the olive fruit fly *Bactrocera oleae*. *Insect Biochem. Mol. Biol.* **33,** 691–699 (2003).
16. Mével-Ninio, M., Terracol, R., Salles, C., Vincent, A. & Payre, F. *ovo*, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with *shavenbaby*, a gene involved in embryo patterning. *Mech. Dev.* **49,** 83–95 (1995).
17. Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genet.* **38,** 594–597 (2006).
18. True, J. R., Mercer, J. M. & Laurie, C. C. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142,** 507–523 (1996).
19. Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287,** 795–801 (1980).
20. Wiellette, E. L. & McGinnis, W. Hox genes differentially regulate *Serrate* to generate segment-specific structures. *Development* **126,** 1985–1995 (1999).
21. Walters, J. W., Munoz, C., Paaby, A. B. & Dinardo, S. Serrate-Notch signaling defines the scope of the initial denticle field by modulating EGFR activation. *Dev. Biol.* **286,** 415–426 (2005).
22. Hatini, V., Green, R. B., Lengyel, J. A., Bray, S. J. & Dinardo, S. The Drumstick/Lines/Bowl regulatory pathway links antagonistic Hedgehog and Wingless signaling inputs to epidermal cell differentiation. *Genes Dev.* **19,** 709–718 (2005).
23. Delon, I., Chanut-Delalande, H. & Payre, F. The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech. Dev.* **120,** 747–758 (2003).
24. Stam, L. F. & Laurie, C. C. Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144,** 1559–1564 (1996).
25. Prud'homme, B. *et al.* Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature* **440,** 1050–1053 (2006).
26. Tao, H., Cox, D. R. & Frazer, K. A. Allele-specific *KRT1* expression is a complex trait. *PLoS Genet.* **2,** e93 (2006).
27. Harr, B., Weiss, S., David, J. R., Brem, G. & Schlötterer, C. A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. *Curr. Biol.* **8,** 1183–1186 (1998).
28. Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, New York, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Tao for *D. mauritiana* stocks and for sharing unpublished data; F. Roch for anti-Min antibody; T. Frankino for help with flies; A. Bassan, P. Valenti, Y. Latapie and S. Plaza for experimental help, discussions and critical reading of the manuscript; and W. Damen and N. Brown for hosting A.P.M. and I.D., respectively, during manuscript revisions. This work was supported by funding from the Association pour la Recherche sur le Cancer (to J.Z.), the Fondation pour la Recherche Médicale (programme équipe 2005), an EMBO Long-Term Fellowship and a BBSRC grant to I.D., a Damon Runyon Cancer Research Foundation Fellowship to V.O., a National Institute of General Medical Sciences National Research Service Award Fellowship to D.G.S., and a NIH grant and a David and Lucile Packard Foundation Fellowship to D.L.S.

Author Contributions The enhancer analysis was designed by A.P.M., I.D., F.P. and D.L.S.; DNA constructs and transgenic flies were made by A.P.M., I.D. and D.G.S.; embryos were stained, examined and photographed by A.P.M., I.D., J.Z., D.G.S., F.P. and D.L.S.; the recombination mapping experiment was designed and performed by D.L.S., V.O. and A.P.M. All authors participated in data analysis and writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.L.S. (dstern@princeton.edu).

METHODS

Reporter constructs. Fragments of the *svb* gene from *D. melanogaster* and *D. sechellia* were amplified by using the Expand PCR system (Roche) (Fig. 2 and Supplementary Table 1). Map position 0 (Fig. 2) corresponds to the *svb* start codon and X chromosome position 4,894,368 of the *D. melanogaster* genome release 4.3. The *svb* transcription start site was mapped to −250 bp by using 5′-rapid amplification of cDNA ends (data not shown). PCR products were cloned into the pCRII vector (Invitrogen) and then sequenced. Fragments were then subcloned into pCaSpeR-hs43-lacZ[29] or pPTGAL[30] and co-injected with pTURBO[31] into *D. melanogaster* w^{1118} embryos under standard conditions[32]. At least three independent transgenic lines were established for each construct and balanced with the use of *w;Sp/CyO;TM6/DrΔ2-3*.

The *D. melanogaster* proximal construct was made as a *Gal-4* construct, and β-galactosidase expression was driven from a *UAS-lacZ* construct. All other constructs were made as *lac-Z* constructs. Thus, the β-galactosidase expression observed for the *D. melanogaster* proximal construct (Fig. 2d, and Supplementary Figs 4k and 5k) may reflect a delay caused by the initial production of Gal-4 protein. This is supported by examination of the mRNA produced by this construct that reveals expression from the *D. melanogaster* proximal enhancer driven in dorsal 4° cells by stage 14 (Supplementary Figs 3c, f and 4c, g).

Enhancer analysis. Expression patterns driven by reporter constructs were assayed by *in situ* hybridizations with antisense *lacZ* or *GAL4* RNA probes labelled with fluorescein or digoxigenin (Roche). To determine the precise expression domains of the enhancer constructs we performed fluorescent double stainings with an anti-β-Gal antibody (Cappel) and either an anti-Min antibody[33], an anti-Engrailed antibody[34] or an anti-Cut antibody[35]. Embryos were examined on Perkin Elmer RS3 Spinning Disk and Leica TSP2 confocal microscopes.

Fine-scale recombination mapping. We first made a *D. mauritiana* chromosome that carried the 4R1 and 3S1 *white*$^{+}$ markers (light and dark orange eyes, respectively) flanking the *svb* gene on a *white*$^{-}$ chromosome. P-element insertions 4R1 and 3S1 correspond to markers 4B(I) and 6BC(I) from ref. 36 and are located in genomic regions AE003432.5 and AE003438.3 (Y. Tao, personal communication), respectively. We crossed these *D. mauritiana* flies with *D. sechellia white*$^{-}$ flies and crossed the F_1 females with *D. simulans* males. We identified female offspring from this cross that carried only one of the two markers on the basis of their eye colour. We then genotyped these recombinants for molecular markers flanking the *svb* locus by using DNA prepared from their wings[37] (Supplementary Table 2). If a female contained a recombination event in the *svb* gene, we crossed her with *D. sechellia* males and scored the resulting progeny for their larval cuticle phenotypes. Half of the larvae were expected to carry the non-recombinant *D. simulans* chromosome and look like *D. simulans*; the other half were expected to carry the recombinant *mauritiana–sechellia* chromosome. Recombination breakpoints within the *svb* gene were mapped by using a further 19 molecular markers (Supplementary Table 2). Larval cuticles were prepared with a 1:1 mixture of Hoyer's medium and lactic acid, and imaged by using phase-contrast or confocal microscopy[38].

29. Thummel, C. S. & Pirrotta, V. New pCaSpeR P element vectors. *Drosophila Info. Serv.* **71,** 150 (1992).
30. Sharma, Y., Cheung, U., Larsen, E. W. & Eberl, D. F. PPTGAL, a convenient Gal4 P-element vector for testing expression of enhancer fragments in *Drosophila. Genesis* **34,** 115–118 (2002).
31. Robertson, H. M. *et al.* A stable genomic source of P element transposase in *Drosophila melanogaster. Genetics* **118,** 461–470 (1988).
32. Rubin, G. M. & Spradling, A. C. Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218,** 348–353 (1982).
33. Roch, F., Alonso, C. R. & Akam, M. *Drosophila miniature* and *dusky* encode ZP proteins required for cytoskeletal reorganisation during wing morphogenesis. *J. Cell Sci.* **116,** 1199–1207 (2003).
34. Patel, N. H. *et al.* Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell* **58,** 955–968 (1989).
35. Blochlinger, K., Bodmer, R., Jan, L. Y. & Jan, Y. N. Patterns of expression of cut, a protein required for external sensory organ development in wild-type and *cut* mutant *Drosophila* embryos. *Genes Dev.* **4,** 1322–1331 (1990).
36. True, J. R., Mercer, J. M. & Laurie, C. C. Differences in crossover frequency and distribution among three sibling species of *Drosophila. Genetics* **142,** 507–523 (1996).
37. Gleason, J. M., Cropp, K. A. & Dewoody, R. S. DNA preparations from fly wings for molecular marker assisted crosses. *Drosophila Info. Serv.* **87,** 107–108 (2004).
38. Stern, D. L. & Sucena, E. in *Drosophila: A Laboratory Manual* (eds Ashburner, M., Hawley, S. & Sullivan, B.) 601–615 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2000).

LETTERS

A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene

Hakon Hakonarson[1,3]*, Struan F. A. Grant[1,3]*, Jonathan P. Bradfield[1]*, Luc Marchand[5], Cecilia E. Kim[1], Joseph T. Glessner[1], Rosemarie Grabs[5], Tracy Casalunovo[1], Shayne P. Taback[6], Edward C. Frackelton[1], Margaret L. Lawson[7], Luke J. Robinson[1], Robert Skraban[1], Yang Lu[5], Rosetta M. Chiavacci[1], Charles A. Stanley[4], Susan E. Kirsch[8], Eric F. Rappaport[9], Jordan S. Orange[10], Dimitri S. Monos[2,10], Marcella Devoto[3,11], Hui-Qi Qu[5] & Constantin Polychronakos[5]

Type 1 diabetes (T1D) in children results from autoimmune destruction of pancreatic beta cells, leading to insufficient production of insulin[1]. A number of genetic determinants of T1D have already been established through candidate gene studies, primarily within the major histocompatibility complex[2–4] but also within other loci[5–12]. To identify new genetic factors that increase the risk of T1D, we performed a genome-wide association study in a large paediatric cohort of European descent. In addition to confirming previously identified loci[2–9], we found that T1D was significantly associated with variation within a 233-kb linkage disequilibrium block on chromosome 16p13. This region contains *KIAA0350*, the gene product of which is predicted to be a sugar-binding, C-type lectin. Three common non-coding variants of the gene (rs2903692, rs725613 and rs17673553) in strong linkage disequilibrium reached genome-wide significance for association with T1D. A subsequent transmission disequilibrium test replication study in an independent cohort confirmed the association. These results indicate that *KIAA0350* might be involved in the pathogenesis of T1D and demonstrate the utility of the genome-wide association approach in the identification of previously unsuspected genetic determinants of complex traits.

The risk for T1D is strongly influenced by multiple genetic loci and environmental factors. The disease is heritable, with first-degree relatives of patients with T1D being at 15-fold greater risk for developing the condition than the general population.

Variation in four loci has already been established to account for a significant proportion of the familial clustering of T1D. These include the major histocompatibility complex (MHC) region on chromosome 6p21 (mostly residing in the *HLA-DRB1*, *-DQA1* and *-DQB1* genes[2–4]), the insulin locus (*INS*) on chromosome 11p15 (refs 5–7), the protein tyrosine phosphatase-22 (*PTPN22*) gene on chromosome 1p13 (refs 8, 9) and the gene that encodes the cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*) on chromosome 2q31 (refs 10–12). The interleukin-2 receptor alpha (*CD25 IL2RA*) locus on chromosome 10p15 (ref. 13) has also been implicated, and a report that T1D is associated with a non-synonymous variant in the innate immunity gene *IFIH1* (ref. 14) remains to be independently replicated. Several other reported associations[15–17] have not been convincingly replicated and remain controversial, and linkage studies[18–21] have established that there is no other locus with an effect size approaching that of the *HLA* genes.

The established genetic associations with T1D explain little more than half of the genetic risk for T1D, indicating that other loci exist, although their number and effect size remain unknown. To search systematically for the remaining loci, we performed a two-stage genome-wide association (GWA) study.

In stage 1, we genotyped 550,000 single nucleotide polymorphisms (SNPs) with the Illumina Human Hap550 Genotyping BeadChip[22], from 563 patients with T1D and 1,146 controls of European ancestry (based on self-report) plus 483 complete T1D family trios of the same ancestry. All patients had clinically proven T1D and were using insulin. We rejected 2,239 SNPs because of call rates $<$90% and 19,065 SNPs were removed for minor allele frequencies below 1%; as such, 534,071 SNPs remained in the analysis. Following this process, 16 trios, 2 cases and 3 controls were removed from further consideration.

In the case-control analysis, we compared single-marker allele frequencies using χ^2 statistics for all markers. We used the transmission disequilibrium test (TDT) to calculate the *P*-values of transmission distortion from heterozygous parents in affected parent–child trios. The resulting *P*-values from the case-control and family-based analyses were then combined using Fisher's method[23] to quantify the overall evidence for association. As anticipated, the MHC region was strongly positive, with 392 markers above the threshold for Bonferroni correction (Supplementary Table S1). As this locus is well established, and as a much denser marker coverage is needed to deal with the particularities of this region, we elected not to address this locus further. However, it should be noted that allele A of the most significant MHC-associated SNP, rs2647044, tags *HLA-DRB1* as efficiently as a previously identified SNP[24] and was observed to be in epistasis ($P < 10^{-10}$) with rs3117098, which is also associated with T1D, at the butyrophilin-like 2 (*BTNL2*) locus within the MHC. We found no other significant epistasis with significantly associated SNPs.

Eleven non-MHC SNPs were the next most significant markers, and remained significant at the 0.05 level after Bonferroni correction (Table 1). One of these eleven markers, rs2476601 ($P = 1.11 \times 10^{-12}$) and another five markers, rs1004446, rs6356, rs10770141, rs7111341

[1]Center for Applied Genomics, and [2]Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [3]Department of Pediatrics and Division of Human Genetics, and [4]Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [5]Departments of Pediatrics and Human Genetics, McGill University, Montreal H3H 1P3, Québec, Canada. [6]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg R3E 0Z2, Manitoba, Canada. [7]Division of Endocrinology, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa K1H 8L1, Ontario, Canada. [8]Markham-Stouffville Hospital, Markham L3P 7P3, Ontario, Canada. [9]The Children's Hospital of Philadelphia Nucleic Acid and Protein Core, Philadelphia, Pennsylvania 19104, USA. [10]Department of Pediatrics, University of Pennsylvania, School of Medicine, Philadelphia, Pennsylvania 19104, USA. [11]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.
*These authors contributed equally to this work.

Table 1 | TDT and case-control association study results for GWA significant markers

		Case-control cohort					Triad cohort ($n = 467$)				
Chr.	SNP	Allele	Aff. allele freq. ($n = 561$)	Ctrl allele freq. ($n = 1{,}143$)	OR (95% CI)	*P*-value	Alleles	Trans:untrans	TDT *P*-value	*P*-value combined	Locus
1	rs2476601	A	0.1471	0.08757	1.80 (1.44, 2.24)	1.32×10^{-7}	A:G	137:64	2.62×10^{-7}	1.11×10^{-12}	*PTPN22*
11	rs1004446	T	0.254	0.3539	0.62 (0.53, 0.73)	4.38×10^{-9}	T:C	160:228	5.56×10^{-4}	6.75×10^{-11}	*INS*
16	rs2903692	A	0.2834	0.3782	0.65 (0.56, 0.76)	4.77×10^{-8}	A:G	170:251	7.89×10^{-5}	1.03×10^{-10}	*KIAA0350*
11	rs6356	A	0.4602	0.3593	1.52 (1.31, 1.76)	1.78×10^{-8}	A:G	255:197	0.00637	2.70×10^{-9}	*INS*
16	rs725613	C	0.3004	0.3898	0.67 (0.58, 0.78)	3.24×10^{-7}	C:A	178:248	6.95×10^{-4}	5.23×10^{-9}	*KIAA0350*
7	rs10255021	A	0.06667	0.1095	0.58 (0.44, 0.77)	1.16×10^{-4}	A:G	18:57	6.69×10^{-6}	1.71×10^{-8}	*COL1A2*
11	rs10770141	A	0.2799	0.373	0.65 (0.56, 0.76)	7.20×10^{-8}	A:G	186:234	0.01917	2.95×10^{-8}	*INS*
1	rs672797	T	0.2257	0.1589	1.54 (1.29, 1.85)	2.67×10^{-6}	T:G	177:119	7.49×10^{-4}	4.20×10^{-8}	*LPHN2*
16	rs17673553	G	0.2023	0.2791	0.66 (0.55, 0.78)	1.30×10^{-6}	G:A	146:203	0.00228	6.12×10^{-8}	*KIAA0350*
11	rs7111341	T	0.1843	0.2631	0.63 (0.53, 0.76)	3.77×10^{-7}	T:C	138:185	0.008919	6.90×10^{-8}	*INS*
11	rs10743152	T	0.271	0.3574	0.67 (0.57, 0.78)	4.73×10^{-7}	T:C	179:233	0.007805	7.53×10^{-8}	*INS*

Minor allele frequencies, *P*-values and odds ratios (OR) are shown. The ORs shown are for the minor alleles (as observed in the controls). Combined *P*-values are also shown, together with the gene in which the markers reside or which they are nearest to. *P*-values are two-sided in each instance. Aff. allele freq., allele frequency in affected individuals; Chr., chromosome; CI, confidence interval; Ctrl allele freq., allele frequency in unaffected individuals; Trans:untrans, ratio of transmitted to untransmitted alleles.

and rs10743152 (*P*-value range 7.53×10^{-8} to 6.75×10^{-11}), are in two known T1D susceptibility loci, *PTPN22* and *INS*, respectively. Three common non-coding variants (rs2903692 allele A, rs725613 allele C and rs17673553 allele G), in strong linkage disequilibrium (LD) in the *KIAA0350* gene on chromosome 16p13.13, also attained genome-wide significance for T1D association (*P*-value range 6.12×10^{-8} to 1.03×10^{-10}; case-control odds ratio (OR) range 0.65–0.66). The minor allele is protective, with a frequency of 0.28–0.39 in controls. In addition, eleven other markers in the *KIAA0350* LD block showed association with $P < 0.00001$ in the family trios and case-control cohort combined (Supplementary Table S2). We found no significant interaction between allele A of rs2903692 and known HLA subtypes (Breslow-Day test for heterogeneity of the allelic odds ratio $P = 0.67$; Supplementary Table S3).

Two other loci, namely the gene for collagen type 1 α2 (*COL1A2*; rs10255021) and rs672797, in the vicinity of latrophilin 2 (*LPHN2*), were also significantly associated following Bonferroni correction in stage 1 (Table 1); however, they failed to replicate in follow-up studies. Thus, in stage 1 we have confirmed the association with the three established T1D loci, and uncovered *KIAA0350* as a potential T1D locus of genome-wide significance. This locus was fast-tracked to Stage 2.

Many reported associations with common variants have not been replicated owing to such factors as population stratification, inadequate statistical power and genotyping errors[25]. Using TDT for the family-based analysis provides adequate protection against stratification, and the clustering of significant SNPs at the *KIAA0350* locus makes genotyping error extremely unlikely; in addition, the application of EIGENSTRAT[26] to the case-control data set indicated that population stratification had little impact on our results (see Supplementary Information). Nevertheless, we sought to confirm the association between T1D and the locus in an additional unrelated sample of affected parent–offspring trios. We used TDT to calculate the level of significance of differences between transmitted and untransmitted allele counts in 1,333 affected offspring from 549 nuclear families from the Type 1 Diabetes Genetics Consortium (T1DGC) plus an additional 390 Canadian trios. Using the SNPlex platform from Sequenom, we confirmed the association of rs17673553, rs725613 and rs2903692 ($P = 0.023 - 0.0022$) and found that several other markers in the LD block also showed association (Supplementary Table S2). All of these SNPs were in LD, with the minor allele again conferring protection, except for the minor A allele of rs7200786, which conferred risk, yielding an OR of 1.33 and a PAR of 12.6% (combined *P* for all three cohorts $= 9.12 \times 10^{-7}$). These SNPs have frequencies in our controls that are very close to those observed in the International HapMap, all are in Hardy–Weinberg equilibrium and all survive all quality control measures for high-quality SNPs.

In an analysis that combined all three independent cohorts (563 cases against 1,146 controls; 483 stage 1 trios and 1,333 T1D offspring from 939 nuclear families for stage 2) for these three intragenic *KIAA0350* markers, the combined *P*-values for their association with T1D ranged from 2.74×10^{-9} to 6.7×10^{-11}. The results remain significant when limited to the 839 nuclear families that self-report as Caucasian (Table 2, Supplementary Table S2).

This locus resides in a 233-kb block of LD that contains only *KIAA0350* and no other genes, making this gene a prime candidate for harbouring the causative variant. *KIAA0350* encodes a protein of unknown function and its genomic location is next to the suppressor of cytokine signalling 1 (*SOCS1*) gene. The almost exclusive expression specificity of *KIAA0350* in immune cells (http://symatlas.gnf.org/SymAtlas), including dendritic cells, B lymphocytes and natural killer (NK) cells, all of which are pivotal in the pathogenesis of T1D[27,28], indicates that the variant probably contributes to the disease by modulating immunity. The predicted protein product of *KIAA0350* bears similarities to a subset of adhesion and immune function signalling molecules. Pfam[29] prediction indicates that this gene probably encodes a protein with a calcium-dependent, or C-type, lectin-binding domain structure. Proteins of this type are known to be involved with calcium current flux, and the predicted function of the protein encoded by *KIAA0350* includes sugar binding, according to the Gene Ontology project (GO: 0005529; http://www.geneontology.org). The C-type lectins are known for their recognition of various carbohydrates and are crucial for processes that range from cell adhesion to pathogen recognition[30].

To investigate whether genotype influences *KIAA0350* expression, we used a synonymous SNP in exon 19, rs2286973, ($r^2 = 0.72$ with rs725613; $r^2 = 0.67$ with rs2903692) to evaluate the relative abundance of each allele in steady-state messenger RNA from ten lymphoblastoid cell lines. The allele ratios in the mRNA, determined through bi-directional sequencing, were not different from those in the DNA, indicating that expression is not influenced by genotype

Table 2 | Replication of stage 1 results for *KIAA0350* in a family-based analysis of an independent cohort derived from 939 nuclear families

		All trios (939 nuclear families)			Caucasians only (839 nuclear families)		Stage 1 and replication
Chr.	SNP	Alleles	Trans: untrans	TDT *P*-value	Trans: untrans	TDT *P*-value	Combined *P*-value
16	rs2903692	A:G	466:538	0.023	438:504	0.032	6.70×10^{-11}
16	rs725613	C:A	461:559	2.15×10^{-3}	435:520	5.95×10^{-3}	8.86×10^{-11}
16	rs17673553	G:A	371:448	7.13×10^{-3}	348:422	7.66×10^{-3}	2.74×10^{-9}

Family-based association *P*-values were computed using TDT. *P*-values are two-sided in each instance.

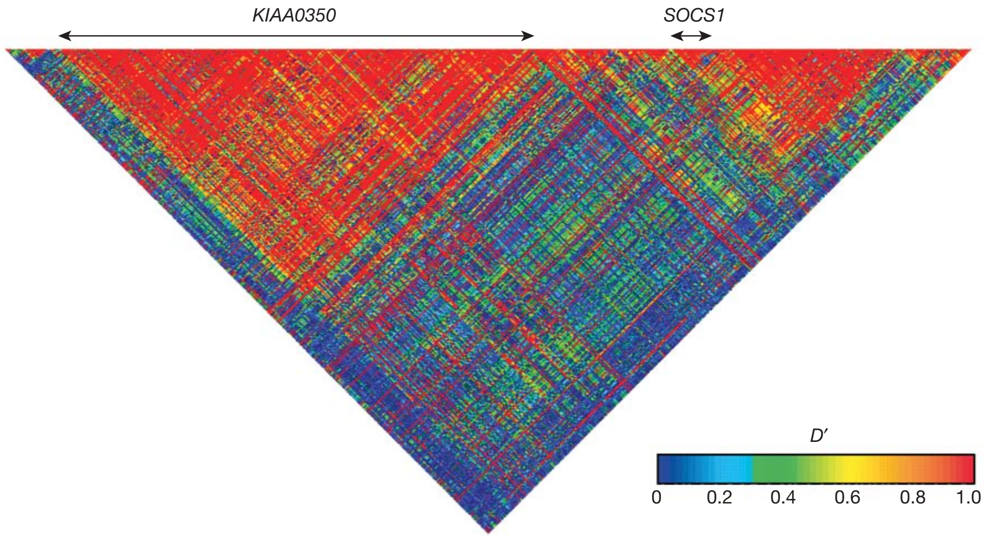


Figure 1 | Pairwise linkage disequilibrium diagram of the *KIAA0350* locus on 16p13.13. This 'gold plot' is derived from HapMap CEPH Utah (CEU) data corresponding to a region spanning from 10,899,122 (rs8063850) to 11,395,501 (rs12597032) base pairs on chromosome 16 (build 35); intensity of shading is proportional to D′ (the fraction of observed linkage disequilibrium over the maximal possible). The relative genomic location of the *KIAA0350* gene is shown; it is contained within a single LD block; no other gene resides within this LD block of association. The most pertinent gene in the adjacent region is the *SOCS1* gene (indicated).

(Supplementary Fig. S1). In an additional attempt to identify the functional variant, we sequenced exon 23 in 20 affected individuals who were homozygous for the risk allele, to determine whether the association might be due to a strong predisposing effect from the rare non-synonymous SNP, rs2241100; however, all individuals were homozygous for the common allele. We also investigated the expression of *KIAA0350* in four different NK cell lines and found a trend towards higher expression in the NKL cell line; interestingly, this cell line is the only one that is homozygous for allele A of rs2903692 (Supplementary Fig. S2).

Studies are underway to characterize the functional role of *KIAA0350*. In light of the crucial role of the MHC genetic repertoire in antigen presentation involving sugar groups, such as lectin, we hypothesize that a genetic variant in the binding site for such a molecule on the activating cytotoxic T-cell could elicit an autoimmune response that results in the destruction of the islet cells of the pancreas, as seen in T1D.

Finally, it is worth noting that haplotype-based coverage with 550-K markers did not reveal loci with effect sizes equal to or stronger than those of *INS* and *PTPN22*. This might lead to the conclusion that such loci, if they exist, cannot be numerous. Although *INS* was unequivocally detected despite the fact that the most associated haplotype at that locus was tagged at an $r^2 < 0.8$, it is still possible that our stage 1 missed loci in that order of effect magnitude because of imperfect tagging. Our approach has a high likelihood of discovering such loci in the course of a full stage 2 and subsequent fine mapping of the loci that will be discovered.

Note added in proof: After the acceptance of this manuscript, an independent GWA was reported that also identified KIAA0350 as a T1D locus[31,32].

METHODS SUMMARY

Cases and family trios for stage 1, as well as 390 of the families that were used in stage 2, were identified through paediatric diabetes clinics in Philadelphia, Montreal, Toronto, Ottawa and Winnipeg; these cases are unique to our study. The remaining stage 2 families were provided by the T1DGC and originated from Europe, North America and Australia. T1DGC families are available to many investigators but we are not aware of any GWA studies that have used them. Controls for stage 1 were drawn from the Children's Hosptial of Philadelphia Health Care Network. Genotyping for stage 1 was conducted using Illumina Infinium Hap500 high-density oligonucleotide microarrays. Genotyping for stage 2 was conducted with matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry using the Sequenom iPLEX system. All statistical tests for association were carried out using the software package *plink*. The single-marker analysis for the genome-wide data was carried out using a χ^2 test on allele count differences in cases and controls. ORs and the corresponding 95% confidence intervals were calculated for the association analysis. A TDT was used to evaluate differences between transmitted and untransmitted allele counts in T1D trios in stage 1 and in nuclear families in stage 2, using the standard TDT implemented in the Haploview software package (http://www.broad.mit.edu/mpg/haploview). The *P*-values from the case-control and family-based analyses in stage 1 were combined using Fisher's method to quantify the overall evidence for association.

On completion of our full stage 2, we will make summary data publicly available (that is, genotype counts for all sets of subjects and transmission counts from heterozygous parents in the family data, at all loci that passed quality control).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 April; accepted 11 June 2007.
Published online 15 July 2007.

1. Devendra, D., Liu, E. & Eisenbarth, G. S. Type 1 diabetes: recent developments. *Br. Med. J.* **328**, 750–754 (2004).
2. Cucca, F. *et al.* A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum. Mol. Genet.* **10**, 2025–2037 (2001).
3. Nerup, J. *et al.* HL-A antigens and diabetes mellitus. *Lancet* **2**, 864–866 (1974).
4. Noble, J. A. *et al.* The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am. J. Hum. Genet.* **59**, 1134–1148 (1996).
5. Bell, G. I., Horita, S. & Karam, J. H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **33**, 176–183 (1984).
6. Bennett, S. T. *et al.* Susceptibility to human type 1 diabetes at *IDDM2* is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature Genet.* **9**, 284–292 (1995).
7. Vafiadis, P. *et al.* Insulin expression in human thymus is modulated by INS VNTR alleles at the *IDDM2* locus. *Nature Genet.* **15**, 289–292 (1997).
8. Bottini, N. *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature Genet.* **36**, 337–338 (2004).
9. Smyth, D. *et al.* Replication of an association between the lymphoid tyrosine phosphatase locus (*LYP/PTPN22*) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes* **53**, 3020–3023 (2004).
10. Kristiansen, O. P., Larsen, Z. M. & Pociot, F. CTLA-4 in autoimmune diseases—a general susceptibility gene to autoimmunity? *Genes Immun.* **1**, 170–184 (2000).

11. Ueda, H. *et al.* Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Nature* **423,** 506–511 (2003).
12. Anjos, S. M., Tessier, M. C. & Polychronakos, C. Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. *J. Clin. Endocrinol. Metab.* **89,** 6257–6265 (2004).
13. Vella, A. *et al.* Localization of a type 1 diabetes locus in the *IL2RA/CD25* region by use of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **76,** 773–779 (2005).
14. Smyth, D. J. *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. *Nature Genet.* **38,** 617–619 (2006).
15. Guo, D. *et al.* A functional variant of SUMO4, a new IκBα modifier, is associated with type 1 diabetes. *Nature Genet.* **36,** 837–841 (2004).
16. Mirel, D. B. *et al.* Association of *IL4R* haplotypes with type 1 diabetes. *Diabetes* **51,** 3336–3341 (2002).
17. Biason-Lauber, A. *et al.* Association of childhood type 1 diabetes mellitus with a variant of PAX4: possible link to beta cell regenerative capacity. *Diabetologia* **48,** 900–905 (2005).
18. Davies, J. L. *et al.* A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* **371,** 130–136 (1994).
19. Concannon, P. *et al.* A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nature Genet.* **19,** 292–296 (1998).
20. Mein, C. A. *et al.* A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nature Genet.* **19,** 297–300 (1998).
21. Cucca, F. *et al.* A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nature Genet.* **19,** 301–302 (1998).
22. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genet.* **37,** 549–554 (2005).
23. Fisher, R. A. *Statistical Methods for Research Workers* edn 13 (Hafner, New York, 1958).
24. de Bakker, P. I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genet.* **38,** 1166–1172 (2006).
25. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4,** 45–61 (2002).
26. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38,** 904–909 (2006).
27. Poirot, L., Benoist, C. & Mathis, D. Natural killer cells distinguish innocuous and destructive forms of pancreatic islet autoimmunity. *Proc. Natl Acad. Sci. USA* **101,** 8102–8107 (2004).
28. Rodacki, M. *et al.* Altered natural killer cells in type 1 diabetic patients. *Diabetes* **56,** 177–185 (2007).
29. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34,** D247–D251 (2006).
30. Cambi, A. & Figdor, C. G. Levels of complexity in pathogen recognition by C-type lectins. *Curr. Opin. Immunol.* **17,** 345–351 (2005).
31. Todd, J. A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature Genet.* advance online publication, doi: 10.1038/ng2068 (6 June 2007).
32. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the use of DNA samples from the T1DGC, funded by the NIH. We thank all participating subjects and families. A. Belisle, A. W. Eckert, A. Estevez, K. Fain, R. Frechette, P. Kline, C. C. Onyiah, G. Otieno, E. Santa, J. L. Shaner, R. M. Smith, A. Thomas and L. Williams helped with genotyping or data collection and management. We thank D. Laforte and the PRUDENT team for subject recruitment and all T1DGC coordinating teams. We also thank S. Kristinsson, L. A. Hermannsson and A. Krisbjörnsson for their software design and contribution. This research was financially supported by the Children's Hospital of Philadelphia, Genome Canada through the Ontario Genomics Institute and the Juvenile Diabetes Research Foundation.

Author Contributions H.H. and C.P. designed the study and supervised the data analysis and interpretation. S.F.A.G., J.P.B. and M.D. conducted the statistical analyses. C.E.K, T.C., E.C.F. and R.S. directed the genotyping of stage 1. H-Q.Q. and C.P. coordinated the genotyping and data analysis for stage 2. Y.L. and H-Q.Q. performed the resequencing and allelic-imbalance experiments. J.S.O. and E.F.R. carried out the work on NKT expression. J.P.B., J.T.G. and L.J.R. provided bioinformatics support. The remaining authors coordinated the studies used in stage 1 and 2. H.H., S.F.A.G., J.P.B., H-Q.Q. and C.P. drafted the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.H. (hakonarson@chop.edu) or C.P. (constantin.polychronakos@mcgill.ca).

METHODS

Type 1 diabetes cohort from Canada. The Canadian cohort consisted of 1,120 nuclear family trios (one affected child and two parents) and 267 independent T1D cases, collected in paediatric diabetes clinics in Montreal, Toronto, Ottawa and Winnipeg. The median age at onset is 8 yr with lower and upper quartiles at 4.6 yr and 11 yr. All patients were diagnosed under the age of 18 and have been treated with insulin since diagnosis, and none has stopped treatment for any reason during that time. Disease diagnosis was based on these clinical criteria, rather than any laboratory tests. Ethnic backgrounds were of mixed European descent, with the largest single subset (409 families) being French Canadian. The Research Ethics Board of the Montreal Children's Hospital and other participating centres approved the study, and written informed consent was obtained from all subjects.

Type 1 Diabetes Genetics Consortium cohort. The Type 1 Diabetes Genetics Consortium cohort consisted of 549 families (2,350 individuals) with at least two children diagnosed with diabetes and both parents available as of the July 2005 data freeze. Criteria were age at diagnosis below 35 yr and had uninterrupted treatment with insulin within six months of diagnosis. For siblings of probands diagnosed under the age of 35 yr, the age-at-diagnosis limit was extended to 45 yr if they were lean and had positive antibodies and/or low C-peptide levels at diagnosis. The median age of onset is 8 yr with quartiles at 4 yr and 13 yr. The samples were collected in Europe, North America and Australia and most subjects were of European ancestry. Autoantibody results are available but were not used to substantiate the diagnosis, except as noted above.

Type 1 diabetes cohort from Philadelphia. The T1D cohort consisted of 103 children recruited at the Children's Hospital of Philadelphia (CHOP) since September 2006. All patients were diagnosed under the age of 18 yr. Of those, 49 T1D patients (32 female, 17 male) were caucasian by self-report (average age of onset 7.07 yr; range 9 months–14 yr) and were included in the analysis. All have been treated with insulin since diagnosis and none has stopped treatment for any reason since. The Research Ethics Board of CHOP approved the study and written informed consent was obtained from all subjects.

Control subjects from Philadelphia. The control group included 1,146 children with self-reported caucasian status, mean age 9.42 yr; 53.05% male and 46.95% female, who did not have diabetes or a first-degree relative with T1D. These individuals were recruited by CHOP's clinicians and nursing staff within the CHOP's Health Care Network, including four primary care clinics and several group practices and outpatient practices that included well child visits. The Research Ethics Board of CHOP approved the study, and written informed consent was obtained from all subjects.

Illumina Infinium assay. We performed high-throughput, genome-wide SNP genotyping, using the InfiniumII HumanHap550 BeadChip technology[22,33] (Illumina), at the Center for Applied Genomics at CHOP. We used 750 ng of genomic DNA to genotype each sample, according to the manufacturer's guidelines. On day one, genomic DNA was amplified 1,000–1,500-fold. On day two, amplified DNA was fragmented to ~300–600 bp, then precipitated and resuspended before being hybridized on to a BeadChip. Single-base extension (SBE) uses a single probe sequence, ~50 bp long, that is designed to hybridize immediately adjacent to the SNP query site. After targeted hybridization to the bead array, the arrayed SNP locus-specific primers (attached to beads) were extended with a single hapten-labelled dideoxynucleotide in the SBE reaction. The haptens were subsequently detected by a multi-layer immunohistochemical sandwich assay, as recently described[22,33]. The Illumina BeadArray Reader scanned each BeadChip at two wavelengths and created an image file. As BeadChip images were collected, intensity values were determined for all instances of each bead type, and data files were created that summarized intensity values for each bead type. These files consisted of intensity data that were loaded directly into Illumina's genotype analysis software, BeadStudio. A bead pool manifest created from the laboratory information management system (LIMS) database containing all the BeadChip data was loaded into BeadStudio along with the intensity data for the samples. BeadStudio used a normalization algorithm to minimize BeadChip to BeadChip variability. Once the normalization was complete, the clustering algorithm was run to evaluate cluster positions for each locus and to assign individual genotypes. Each locus was given an overall score, which was based on the quality of the clustering, and each individual genotype call was given a GenCall score. GenCall scores provided a quality metric that ranges from 0 to 1 assigned to every genotype called. GenCall scores were then calculated using information from the clustering of the samples. The location of each genotype relative to its assigned cluster determined its GenCall score.

Sequenom iPLEX assay. Genotypes for the rapid confirmation study were obtained using the iPLEX assay (Sequenom). Locus-specific PCR primers and allele-specific detection primers were designed using the MassARRAY Assay Design 3.0 software (Sequenom). The sample DNAs were amplified in a 34-plex PCR reaction and labelled using a locus-specific single-base extension reaction. The resulting products were desalted and transferred to a 384-element SpectroCHIP array. Allele detection was performed using MALDI-TOF MS. The mass spectrograms were analysed by the MassARRAY TYPER software (Sequenom). The 90 CEU (European-descent individuals genotyped in HapMap) were included as accuracy controls.

Statistical approaches. All statistical tests for association were carried out using the software package *plink* (http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml)[34]. The single marker analysis for the genome-wide data was carried out using a χ^2 test on allele count differences between 563 cases and 1,146 controls. Odds ratios and the corresponding 95% confidence intervals were calculated for the association analysis. The TDT was used to calculate *P*-values on differences between transmitted and untransmitted allele counts in 467 T1D trios in Stage 1 and in 939 nuclear families in Stage 2. Counts of untransmitted and transmitted alleles from heterozygous parents to affected offspring were determined using the standard TDT implemented in the Haploview software package (;http://www.broad.mit.edu/mpg/haploview)[35]. The *P*-values from the case-control and family-based analyses in stage 1 were combined using Fisher's method[23] to quantify the overall evidence for association.

Because some of our samples (mainly the controls) were not genotyped for HLA-DQB1, tags were chosen on the basis of the Wellcome Trust Sanger Institute data (http://www.sanger.ac.uk/HGP/Chr6/ng2006-data). One SNP was chosen to tag DQB1*0201, namely rs2187668, at $r^2 = 0.51$. A two-marker haplotype was used to tag DQB1*0302, namely rs9275184 and rs241448, at $r^2 = 0.96$. To test how well these tags worked to predict HLA status with respect to these two surrogates, the subset of samples that were HLA typed were tested, yielding a correct predication rate of 82.7%. The remainder of the sample was then predicted on the basis of these selected tags and thus the stratification analysis was carried out using these predictions.

33. Steemers, F. J. *et al.* Whole-genome genotyping with the single-base extension assay. *Nature Methods* 3, 31–33 (2006).
34. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* (in the press).
35. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265 (2005).

LETTERS

Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer

Scott A. Tomlins[1]*, Bharathi Laxman[1]*, Saravana M. Dhanasekaran[1]*, Beth E. Helgeson[1], Xuhong Cao[1], David S. Morris[2], Anjana Menon[1], Xiaojun Jing[1], Qi Cao[1], Bo Han[1], Jindan Yu[1], Lei Wang[1], James E. Montie[2,4], Mark A. Rubin[5,6], Kenneth J. Pienta[2,3,4], Diane Roulston[1], Rajal B. Shah[1,2,4], Sooryanarayana Varambally[1,4], Rohit Mehra[1,4] & Arul M. Chinnaiyan[1,2,4]

Recently, we identified recurrent gene fusions involving the 5′ untranslated region of the androgen-regulated gene *TMPRSS2* and the ETS (E26 transformation-specific) family genes *ERG*, *ETV1* or *ETV4* in most prostate cancers[1,2]. Whereas *TMPRSS2–ERG* fusions are predominant, fewer *TMPRSS2–ETV1* cases have been identified than expected on the basis of the frequency of high (outlier) expression of *ETV1* (refs 3–13). Here we explore the mechanism of *ETV1* outlier expression in human prostate tumours and prostate cancer cell lines. We identified previously unknown 5′ fusion partners in prostate tumours with *ETV1* outlier expression, including untranslated regions from a prostate-specific androgen-induced gene (*SLC45A3*) and an endogenous retroviral element (*HERV-K_22q11.23*), a prostate-specific androgen-repressed gene (*C15orf21*), and a strongly expressed housekeeping gene (*HNRPA2B1*). To study aberrant activation of *ETV1*, we identified two prostate cancer cell lines, LNCaP and MDA-PCa 2B, that had *ETV1* outlier expression. Through distinct mechanisms, the entire *ETV1* locus (7p21) is rearranged to a 1.5-megabase prostate-specific region at 14q13.3–14q21.1 in both LNCaP cells (cryptic insertion) and MDA-PCa 2B cells (balanced translocation). Because the common factor of these rearrangements is aberrant *ETV1* overexpression, we recapitulated this event *in vitro* and *in vivo*, demonstrating that *ETV1* overexpression in benign prostate cells and in the mouse prostate confers neoplastic phenotypes. Identification of distinct classes of ETS gene rearrangements demonstrates that dormant oncogenes can be activated in prostate cancer by juxtaposition to tissue-specific or ubiquitously active genomic loci. Subversion of active genomic regulatory elements may serve as a more generalized mechanism for carcinoma development. Furthermore, the identification of androgen-repressed and insensitive 5′ fusion partners may have implications for the anti-androgen treatment of advanced prostate cancer.

Recurrent chromosomal rearrangements have been causally implicated in haematological and mesenchymal malignancies; although predicted to occur in common epithelial carcinomas, they have not been well characterized[14,15]. Using a bioinformatics strategy to nominate genes showing high (outlier) expression in a subset of cancers, we identified fusions of the 5′-untranslated region of *TMPRSS2* (21q22) to *ERG* (21q22), *ETV1* (7p21) or *ETV4* (17q21) in cases that overexpressed the respective ETS family member[1,2]. *TMPRSS2* had previously been characterized as androgen-regulated[16], and its androgen-responsive regulatory elements drive ETS family member outlier expression[1,17]. Thus, fusions between *TMPRSS2* and ETS family members are functionally similar to haematological malignancy rearrangements in which tissue-specific promoter or enhancer elements of one gene are juxtaposed to proto-oncogenes[15,18].

Multiple studies have confirmed the presence of *TMPRSS2–ERG* fusions in 36–78% of prostate cancers from prostate-specific-antigen-screened surgical cohorts (Supplementary Table 1). Approximately 90% of samples with *ERG* outlier expression harbour *TMPRSS2–ERG* fusions[1,19], confirming this as the predominant mechanism driving *ERG* overexpression. In contrast, although microarray studies show *ETV1* outlier expression in 6–16% of prostate cancer samples, only 2 of 205 (1.0%) samples analysed harboured *TMPRSS2–ETV1* fusions (Supplementary Table 1).

Here we addressed this discrepancy between *ETV1* outlier and *TMPRSS2–ETV1* frequencies. By quantitative PCR (Q-PCR) across 2 cohorts, 26 and 3 of 54 localized prostate cancer samples showed *ERG* (48%) and *ETV1* (5.5%) outlier expression, respectively (Supplementary Fig. 1). Additionally, two hormone-refractory metastatic prostate cancer samples, MET26 (our *TMPRSS2–ETV1* index case[1]) and MET23, showed *ETV1* outlier expression. However, other than MET26, no samples expressed *TMPRSS2–ETV1* fusion transcripts.

To characterize the *ETV1* transcript in outlier cases, we performed 5′-RNA-ligase-mediated rapid amplification of complementary DNA ends (RLM-RACE). Instead of 5′ exons from *TMPRSS2*, the other four samples contained unique 5′ sequences (Fig. 1a). In PCa_ETV1_1, exons 1–4 of *ETV1* were replaced with two exons from 22q11.23 that had homology to human endogenous retrovirus family K (referred to as *HERV-K_22q11.23*). In PCa_ETV1_2, exon 1 of *ETV1* was replaced with exon 1 of *HNRPA2B1* (7p15), whereas in PCa_ETV1_3, exons 1–4 of *ETV1* were replaced with a 5′-extended exon 1 of *SLC45A3* (1q32). In MET23, exons 1–5 of *ETV1* were replaced with exons 1–2 from *C15orf21* (15q21) (Fig. 1a). We confirmed these fusion transcripts by Q-PCR and genomic fusions by fluorescent *in situ* hybridization (FISH; Supplementary Figs 2 and 3). Additional information about the 5′ partners and the FISH results is described in Supplementary Discussion.

HERV-K_22q11.23–ETV1, *SLC45A3–ETV1* and *C15orf21–ETV1* fusions contain no predicted translated sequences from the 5′ partner, and *HNRPA2B1* would only contribute two residues to a *HNRPA2B1–ETV1* fusion protein. Because their regulatory elements probably drive aberrant *ETV1* expression, we characterized the tissue specificity and androgen regulation of these 5′ partners by microarray or massively parallel signature sequencing (MPSS, as described

[1]Michigan Center for Translational Pathology, Department of Pathology, [2]Department of Urology, [3]Department of Internal Medicine, and [4]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [5]Department of Pathology, Brigham and Women's Hospital, [6]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

in the Methods). Similar to *TMPRSS2*, *SLC45A3* showed marked overexpression in prostate cancer (median is 2.45 standard deviations above the median value per array) compared to other tumour types (median = 0.33, $P = 2.4 \times 10^{-7}$) in a large DNA microarray study. *C15orf21* showed similar overexpression in prostate cancer ($P = 3.4 \times 10^{-6}$). In contrast, *HNRPA2B1* showed high expression in prostate and other tumour types (median = 2.36 versus 2.41, $P > 0.05$) (Fig. 1b). By MPSS, *HERV-K_22q11.23* was shown to be highly expressed in normal prostate (94 transcripts per million) compared to the 31 other normal tissues (median = 9 transcripts per million; Fig. 1b). By Q-PCR, endogenous expression of *SLC45A3* (21.6-fold, $P = 6.5 \times 10^{-4}$) and *HERV-K_22q11.23* (7.8-fold, $P = 2.4 \times 10^{-4}$) in the LNCaP prostate cancer cell line was shown to be greatly increased by the synthetic androgen R1881, similar to *TMPRSS2* (14.8-fold, $P = 9.95 \times 10^{-7}$). Conversely, the expression of *C15orf21* was significantly decreased (1.9-fold, $P = 0.0012$) and the expression of *HNRPA2B1* was not affected by R1881 stimulation (1.17-fold, $P = 0.29$) (Fig. 1c).

We next sought to identify cell-line models of *ETV1* outlier expression. Previously, we reported that the LNCaP cell line markedly overexpressed *ETV1*, however RLM-RACE revealed expression of only the wild-type transcript[1]. We proposed that LNCaP may harbour a previously unknown rearrangement affecting the expression of *ETV1*, and used a split-probe FISH strategy to look for gross rearrangements (Fig. 2). On LNCaP metaphase spreads, this assay revealed two pairs of co-localizing signals at the *ETV1* locus (7p) and two split signals in which the 5′ signals remained on 7p whereas the 3′ probes (overlapping the *ETV1* locus) were inserted into another chromosome (Fig. 2b). As described in the Supplementary Discussion and Supplementary Figs 4–8, we identified this rearrangement as a cryptic insertion of a minimal region around *ETV1* into an intronic sequence from the *MIPOL1* locus at 14q13.3–14q21.1 in the LNCaP cell line (Fig. 2d).

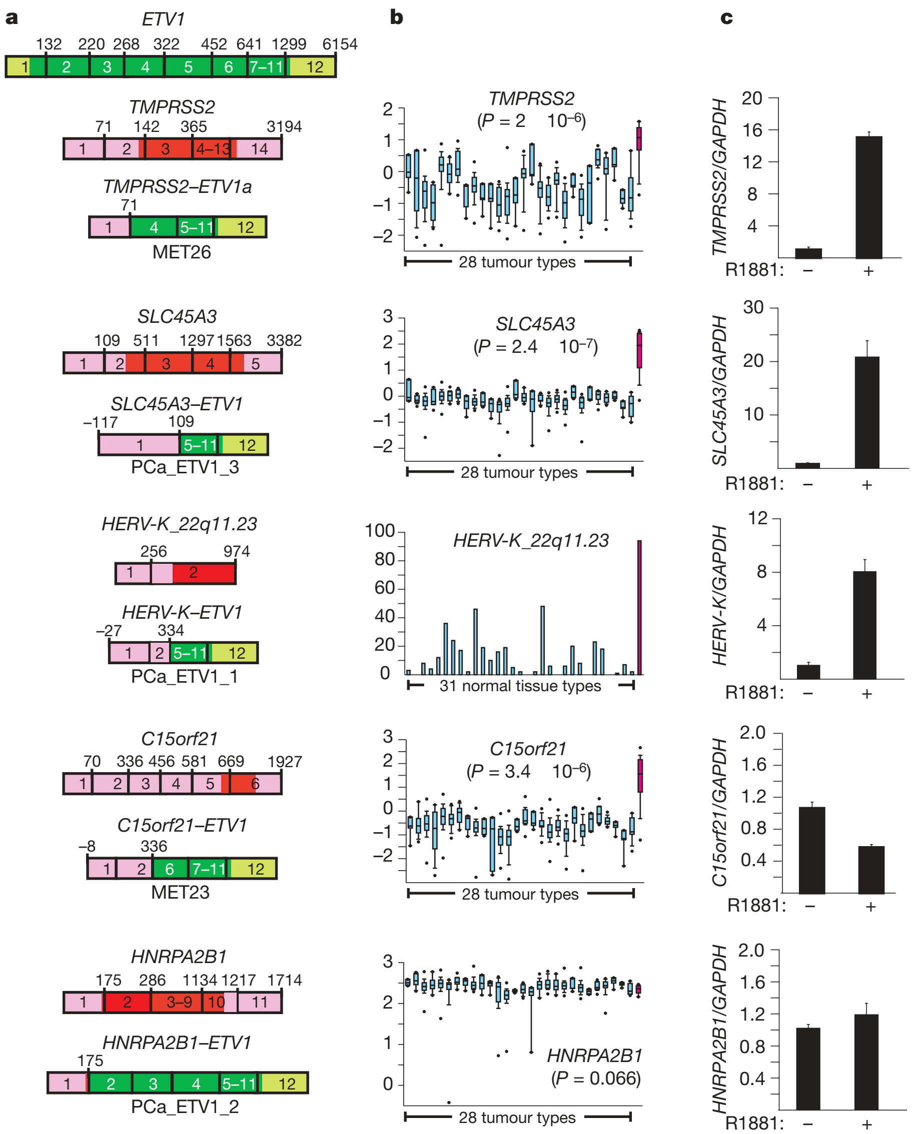


Figure 1 | Identification of prostate-specific or ubiquitously active regulatory elements fused to *ETV1*. **a**, Structure of new 5′ partners fused to *ETV1* in outlier cases. Structures of *ETV1* and all 5′ partners are based on sequences listed in Supplementary Table 2. The numbers in the boxes represent exons. The numbers above the boxes indicate the last base of each exon. Untranslated regions are in lighter shades (pink and light green). **b**, Tissue specificity of 5′ fusion partners was determined in normal tissues or cancers (blue) and in normal prostate or prostate cancer (magenta) (see Methods). See Supplementary Table 4 for tissue and tumour classes. The significance of prostate cancer versus all other tumours is indicated. Box and whisker plots show median +/− 90th%/10th%. **c**, Assessment of androgen regulation of 5′ fusion partners. Endogenous expression of 5′ fusion partners (normalized to the housekeeping gene *GAPDH*) was assessed by Q-PCR in LNCaP cells with (+) or without (–) stimulation by the synthetic androgen R1881 (mean ($n = 4$) plus s.e.).

By screening additional prostate cancer cell lines for *ETV1* overexpression, we identified *ETV1* outlier expression in MDA-PCa 2B (Supplementary Fig. 9). A previous analysis of MDA-PCa 2B demonstrated the presence of a balanced t(7;14)(p21;q21)[20] translocation corresponding to the locations of the *ETV1* and *MIPOL1* loci. We demonstrate that MDA-PCa 2B also harbour a rearrangement involving *ETV1*, because the *ETV1* locus translocates to the d14 (Fig. 2c). The 1.5-Mb 14q13.3–14q21.1 region is the partner of this balanced translocation, because the telomeric 14q13.3–14q21.1 probe localizes to the d7 (Supplementary Fig. 7).

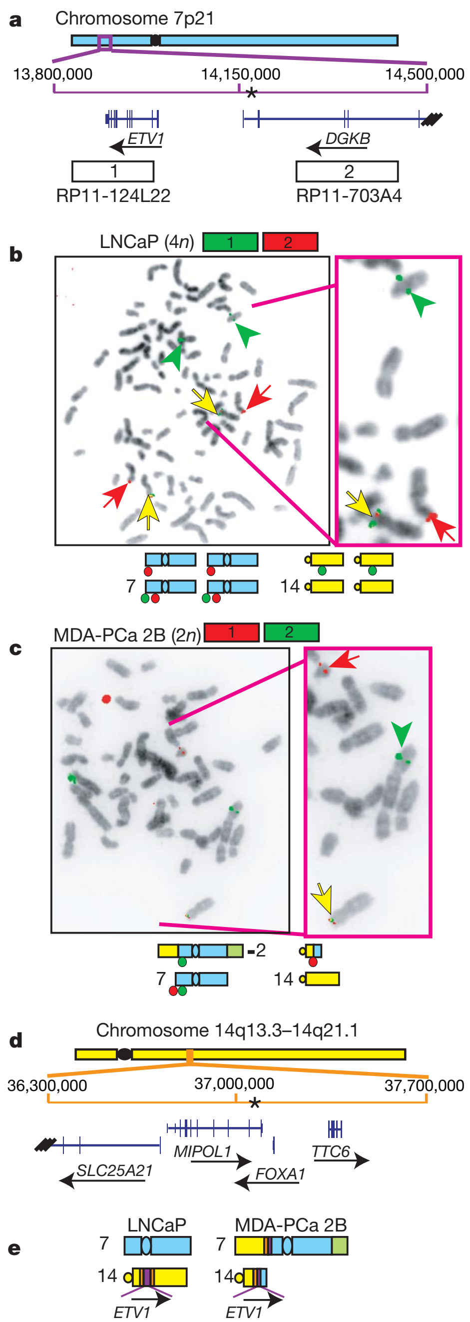


Figure 2 | *ETV1* is rearranged to 14q13.3–14q21.1 in LNCaP and MDA-PCa 2B. a, Schematic showing the *ETV1* locus (purple) on chromosome 7 (blue) and the BACs (rectangles) used for FISH (adapted from the UCSC Genome Browser). Genes are shown and the direction of transcription is indicated by arrows. The LNCaP breakpoint (see Supplementary Figs 4–8) is indicated by an asterisk. **b**, **c**, FISH using the indicated BACs on LNCaP (**b**, tetraploid) and MDA-PCa 2B (**c**, diploid) metaphase spreads. Co-localized signals are indicated by yellow arrows, and separate red and green signals (indicating rearrangements) are indicated by red arrows and green arrowheads, respectively. Schematics of probe localization and chromosome structures as determined by spectral karyotyping[20,28] are indicated. **d**, Schematic of 14q13.3–14q21.1 (orange) on chromosome 14 (yellow). **e**, Structure of *ETV1* and 14q13.3–14q21.1 in LNCaP and MDA-PCa 2B.

The existence of mechanistically distinct rearrangements resulting in the localization of *ETV1* to 14q13.3–14q21.1 (Fig. 2e) in prostate cancer cell lines with *ETV1* outlier expression suggests that elements in this region mediate the aberrant expression of *ETV1*. By characterizing the tissue specificity and androgen regulation of the four contiguous genes at the 14q13.3–14q21.1 breakpoint (Fig. 2d; *SLC25A21*, *MIPOL1*, *FOXA1* and *TTC6*) as well as that of *ETV1* in LNCaP and its androgen-insensitive derivative C4-2B, we demonstrate that this region is both prostate-specific and coordinately regulated by androgen (Supplementary Figs 10–12 and Supplementary Discussion).

The 5′ partners do not contribute a coding sequence to the *ETV1* transcript, therefore the common result of the different *ETV1* rearrangements in clinical samples and prostate cancer cell lines is aberrant overexpression of truncated *ETV1*. We recapitulated this event *in vitro* and *in vivo* to determine the role of aberrant ETS family member expression in prostate cancer. We designed adenoviral and lentiviral constructs to overexpress *ETV1*, as expressed in our index *TMPRSS2–ETV1* fusion-positive case MET26 (Supplementary Fig. 13a). In RWPE and PrEC cells, *ETV1* overexpression had no detectable effect on proliferation (Supplementary Fig. 13b, c). *ETV1* overexpression had no effect on the percentage of RWPE cells in the S phase of the cell cycle (Supplementary Fig. 13d) and was not sufficient for transformation (Supplementary Fig. 13e). However, *ETV1* overexpression markedly increased invasion in a modified basement membrane invasion assay in RWPE (3.4-fold, $P = 0.0005$) (Fig. 3a) and PrEC (6.3-fold, $P = 0.0006$) (Supplementary Fig. 14a). Additionally, *ETV1* knockdown in LNCaP using either short interfering (si)RNA or short hairpin (sh)RNA inhibited invasion (Fig. 3b and Supplementary Fig. 14b–d), consistent with previous work[21]. To investigate the transcriptional programme regulated by *ETV1*, we profiled stable RWPE-*ETV1* cells and analysed the expression signatures using the Oncomine Concepts Map (OCM, http://www.oncomine.org). The OCM is a resource to look for associations between more than 20,000 biologically related gene sets by disproportionate overlap[10,22]. As shown in Fig. 3c, OCM analysis identified a network of molecular concepts related to cell invasion that were enriched in our *ETV1*-overexpressed signature, consistent with the phenotypic effects described above. Specific examples are described in Supplementary Discussion.

We next determined the effects of *ETV1* overexpression *in vivo*. We generated transgenic mice expressing a Flag-tagged, truncated version of *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*) (Supplementary Fig. 13a), which drives strong transgene expression exclusively in the prostate under androgen regulation[23]. This transgene is functionally analogous to the androgen-induced gene fusions of *ETV1* we identified in human prostate cancer. By 12–14 weeks of age, 6 of 8 (75%) ARR2Pb-*ETV1* mice developed mouse prostatic intraepithelial neoplasia (mPIN) (Fig. 4 and Supplementary Table 5). Consistent with the definition of mPIN[24], we observed focal proliferative lesions contained within normal glands in the prostates of ARR2Pb-*ETV1* mice (Fig. 4a–d); these lesions exhibited nuclear atypia, including stratification, hyperchromasia and macronucleoli. mPIN was observed in all three prostatic lobes (anterior, ventral and dorsolateral) of ARR2Pb-*ETV1* mice, most commonly in the ventral lobe (7 of 11, 63.6%) (Supplementary Table 5). By immunohistochemistry in ARR2Pb-*ETV1* mice, we observed strong ETV1–Flag expression exclusively in mPIN foci and not in benign glands (data not shown). Although we have not observed progression to invasive carcinoma in ARR2Pb-*ETV1* mice, we have only characterized 4 mice older than 19 weeks of age, 3 of which (75%) also had mPIN (Supplementary Table 5); this suggests that additional genetic lesions are required for the development of carcinoma. Combined with our *in vitro* observations, these

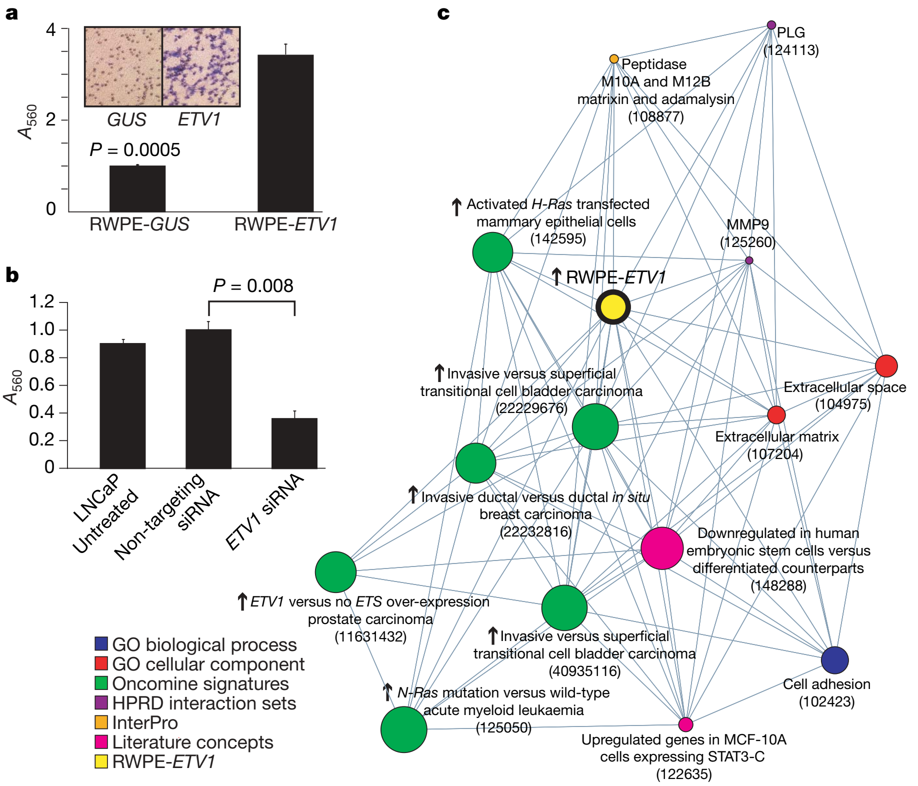


Figure 3 | *ETV1* overexpression in prostate cells confers invasiveness. **a**, We infected the benign prostate cell line RWPE with lentiviruses expressing *ETV1* (exon 4 to the reported stop codon) or control (*GUS*). Stable clones were assayed for invasion through a modified basement membrane (mean ($n = 3$) plus s.e.). Photomicrographs of invaded cells are shown. **b**, LNCaP cells were treated with transfection reagent alone (untreated) or were transfected with non-targeting or *ETV1* siRNA, and assessed for invasion as in **a** (mean ($n = 3$) plus s.e.). **c**, Oncomine concepts map of genes overexpressed in RWPE-*ETV1* compared to RWPE-*GUS* cells (yellow node). Nodes represent molecular concepts (biologically related gene sets). Node size is proportional to the number of genes in the concept, and the colour indicates the concept type. Each edge represents a significant enrichment ($P < 0.005$). HPRD interaction sets for MMP9 and PLG represent known interactors for each gene product. Oncomine concept map IDs are provided for all concepts.

results demonstrate that *ETV1* induces a neoplastic phenotype in the mouse prostate and supports an oncogenic role for ETS gene fusions in human prostate cancer.

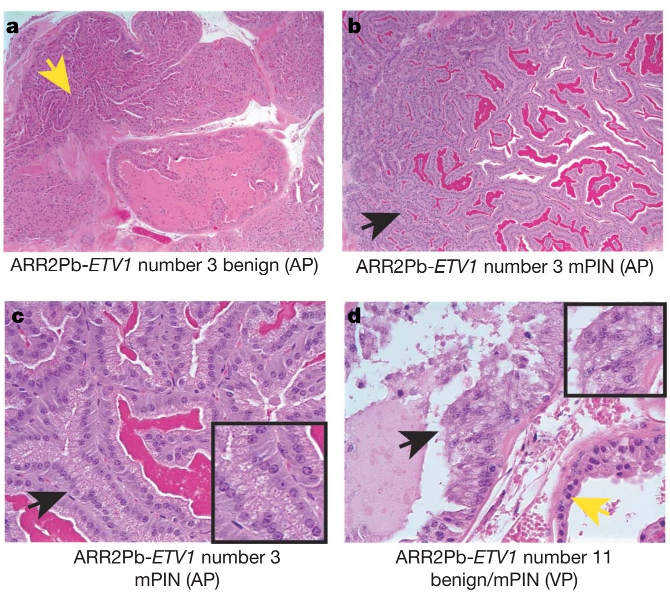


Figure 4 | Transgenic mice expressing *ETV1* develop mPIN. We generated transgenic mice expressing *ETV1* under the control of the modified probasin promoter (ARR2Pb-*ETV1*) (Supplementary Fig. 13a). Mice were killed at 12–43 weeks and mPIN was observed in 75% of ARR2Pb-*ETV1* mice (Supplementary Table 5). **a–d**, Haematoxylin-and eosin-stained ARR2Pb-*ETV1* prostates. Benign epithelia and areas of mPIN are indicated by yellow and black arrows, respectively. Consistent with the definition of mPIN, normal areas (**a**) and mPIN (**b**) were observed in the anterior prostate (AP) of an ARR2Pb-*ETV1* mouse. **c**, High-power view of **b**. **d**, Normal glands and mPIN foci in the ventral prostate (VP) of an ARR2Pb-*ETV1* mouse. Original magnification for **a** and **b** is ×100, and for **c** and **d** is ×400. Insets in **c** and **d** show prominent macronucleoli.

Including fusions between *TMPRSS2* and ETS family members, we have now identified five classes of ETS rearrangements in prostate cancer (Supplementary Fig. 15). Identification of untranslated regions from the prostate-specific androgen-induced gene *TMPRSS2* has provided a mechanism for aberrant ETS family member expression. Thus, fusions between *TMPRSS2* and ETS family members (class I) represent the predominant class of ETS rearrangements in prostate cancer. Rearrangements involving fusions with untranslated regions from other prostate-specific androgen-induced 5′ partner genes (class IIa) or endogenous retroviral elements (class IIb) are probably functionally similar to *TMPRSS2*–ETS rearrangements. Similar to 5′ partners in class I and II ETS rearrangements, *C15orf21* is markedly overexpressed in prostate cancer; however, because *C15orf21* is repressed by androgen, this represents a previously unknown class of rearrangements (class III) involving prostate-specific androgen-repressed 5′ partners.

In contrast, *HNRPA2B1*, which encodes a member of the ubiquitously expressed heteronuclear ribonuclear proteins, did not show prostate-specific expression or androgen responsiveness. Thus, *HNRPA2B1*–*ETV1* represents a previously unknown class of ETS rearrangements (class IV), in which non-tissue-specific promoter elements drive ETS expression. Whereas class I–III ETS rearrangements are functionally analogous to *IGH*–*MYC* rearrangements in B-cell malignancies, *HNRPA2B1*–*ETV1* is more analogous to inv(3)(q21q26) and t(3;3)(q21;q26) in acute myeloid leukaemia, which are thought to place *EVI1* (ecotropic viral integration site 1) under the control of enhancer elements of the constitutively expressed *RPN1* gene[25,26].

By screening prostate cancer cell lines that have *ETV1* outlier expression, we identified rearrangements in LNCaP and MDA-PCa 2B that result in the localization of *ETV1* to 14q13.3–14q21.1. Because this aberration is recurrent in prostate cancer cell lines, we propose that characterizing additional prostate cancer cases with *ETV1* outlier expression will identify clinical specimens with similar rearrangements (class V), in which the entire ETS family gene is rearranged to prostate-specific regions. The identification of distinct

classes of 5′ fusion partners has implications for the detection of gene fusions in prostate cancer and may be important for management, particularly with regard to the effects of androgen ablation on the expression of the different ETS rearrangement classes (Supplementary Fig. 16), as described in Supplementary Discussion.

Multiple classes of gene rearrangements in prostate cancer suggest a generalized role for chromosomal rearrangements in common epithelial cancers. For example, tissue-specific promoter elements may be fused to oncogenes in other hormone-driven cancers, such as breast cancer. Additionally, whereas prostate-specific fusions would not provide a growth advantage and be selected for in other epithelial cancers, fusions involving strong promoters of ubiquitously expressed genes, such as *HNRPA2B1*, could result in the aberrant expression of oncogenes across tumour types. This study supports a role for chromosomal rearrangements in common epithelial tumour development through a variety of mechanisms, similar to haematological malignancies.

METHODS SUMMARY

Q-PCR, RLM-RACE for *ETV1* fusions, androgen stimulation of LNCaP cells and interphase FISH were performed essentially as described[1,2] using indicated oligonucleotide primers (Supplementary Table 2) and bacterial artificial chromosome (BAC) probes (Supplementary Table 3). Tissue-specific expression of 5′ fusion partners was determined using the International Genomics Consortium's expO data set accessed in the Oncomine database[27] and the Lynx Therapeutics MPSS data set (GSE1747). Expression profiling was performed using Agilent Whole Human Genome Oligo Microarrays. Adenoviruses and lentiviruses expressing *ETV1* were generated by the University of Michigan Vector Core. Transgenic ARR2Pb-*ETV1* mice were generated by the University of Michigan Transgenic Animal Model Core.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 April; accepted 16 June 2007.

1. Tomlins, S. A. *et al.* Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. *Science* **310,** 644–648 (2005).
2. Tomlins, S. A. *et al. TMPRSS2:ETV4* gene fusions define a third molecular subtype of prostate cancer. *Cancer Res.* **66,** 3396–3400 (2006).
3. Cerveira, N. *et al. TMPRSS2–ERG* gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. *Neoplasia* **8,** 826–832 (2006).
4. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113,** 913–923 (2004).
5. Hermans, K. G. *et al. TMPRSS2:ERG* fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. *Cancer Res.* **66,** 10658–10663 (2006).
6. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101,** 811–816 (2004).
7. Mehra, R. *et al.* Comprehensive assessment of *TMPRSS2* and *ETS* family gene aberrations in clinically localized prostate cancer. *Mod. Pathol.* **6,** 1177–1187 (2007).
8. Perner, S. *et al. TMPRSS2:ERG* fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res.* **66,** 8337–8341 (2006).
9. Soller, M. J. *et al.* Confirmation of the high frequency of the *TMPRSS2/ERG* fusion gene in prostate cancer. *Genes Chromosom. Cancer* **45,** 717–719 (2006).
10. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nature Genet.* **39,** 41–51 (2007).
11. Winnes, M., Lissbrant, E., Damber, J. E. & Stenman, G. Molecular genetic analyses of the *TMPRSS2–ERG* and *TMPRSS2–ETV1* gene fusions in 50 cases of prostate cancer. *Oncol. Rep.* **17,** 1033–1036 (2007).
12. Yoshimoto, M. *et al.* Three-color FISH analysis of *TMPRSS2/ERG* fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. *Neoplasia* **8,** 465–469 (2006).
13. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22,** 2790–2799 (2004).
14. Mitelman, F., Johansson, B. & Mertens, F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nature Genet.* **36,** 331–334 (2004).
15. Rowley, J. D. Chromosome translocations: dangerous liaisons revisited. *Nature Rev. Cancer* **1,** 245–250 (2001).
16. Lin, B. *et al.* Prostate-localized and androgen-regulated expression of the membrane-bound serine protease *TMPRSS2*. *Cancer Res.* **59,** 4180–4184 (1999).
17. Mertz, K. D. *et al.* Molecular characterization of the *TMPRSS2–ERG* gene fusion in the NCI-H660 prostate cancer cell line—a new perspective for an old model. *Neoplasia* **9,** 200–206 (2007).
18. Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature* **372,** 143–149 (1994).
19. Demichelis, F. *et al. TMPRSS2:ERG* gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene* advance online publication, doi: 10.1038/sj.onc.1210237 (22 January 2007).
20. van Bokhoven, A. *et al.* Spectral karyotype (SKY) analysis of human prostate carcinoma cell lines. *Prostate* **57,** 226–244 (2003).
21. Cai, C. *et al. ETV1* is a novel androgen receptor-regulated gene that mediates prostate cancer cell invasion. *Mol. Endocrinol.* advance online publication, doi:10.1210/me.2006-0480 (15 May 2007).
22. Rhodes, D. R. *et al.* Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9,** 443–454 (2007).
23. Ellwood-Yen, K. *et al.* Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* **4,** 223–238 (2003).
24. Shappell, S. B. *et al.* Prostate pathology of genetically engineered mice: definitions and classification. The consensus report from the Bar Harbor meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee. *Cancer Res.* **64,** 2270–2305 (2004).
25. Suzukawa, K. *et al.* Identification of a breakpoint cluster region 3′ of the ribophorin I gene at 3q21 associated with the transcriptional activation of the EVI1 gene in acute myelogenous leukemias with inv(3)(q21q26). *Blood* **84,** 2681–2688 (1994).
26. Wieser, R. Rearrangements of chromosome band 3q21 in myeloid leukemia. *Leuk. Lymphoma* **43,** 59–65 (2002).
27. Rhodes, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9,** 166–180 (2007).
28. Beheshti, B., Karaskova, J., Park, P. C., Squire, J. A. & Beatty, B. G. Identification of a high frequency of chromosomal rearrangements in the centromeric regions of prostate cancer cell lines by sequential giemsa banding and spectral karyotyping. *Mol. Diagn.* **5,** 23–32 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Rhodes, S. Kalyana-Sundaram and T. Barrette for support of the OCM, L. Smith for cytogenetics assistance, E. Keller and J. Macoska for prostate cancer cell lines, J. Moran for discussions regarding endogenous retroviral elements, J. Presner for technical assistance, and R. Craig and L. Stoolman for the FACS analysis. We thank the UM Transgenic Animal Model Core for generating transgenic mice and the UM Vector core for virus generation. This work was supported in part by Department of Defense (to R.M., A.M.C. and S.V.), the National Institutes of Health (to K.J.P., A.M.C., R.B.S., K.J.P. and A.M.C.), the Early Detection Research Network (to A.M.C.), the Prostate Cancer Foundation (to A.M.C.), and Gen-Probe Incorporated (to A.M.C.). A.M.C. is supported by a Clinical Translational Research Award from the Burroughs Wellcome Foundation. S.A.T. is supported by a Rackham Predoctoral Fellowship. K.J.P. is supported as an American Cancer Society Clinical Research Professor. S.A.T. is a Fellow of the Medical Scientist Training Program.

Author Information The primary microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) under the GEO series accession numbers GSE7701 and GSE7702. Sequences of the *ETV1* fusion transcript junctions identified by RACE have been deposited in GenBank under accession numbers EF632109–EF632112. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M.C. (arul@umich.edu).

METHODS

Samples and cell lines. Prostate tissues were from the radical prostatectomy series at the University of Michigan and from the Rapid Autopsy Program[29], both of which are part of University of Michigan Prostate Cancer Specialized Program of Research Excellence Tissue Core. All samples were collected with informed consent of the patients and previous institutional review board approval.

The benign immortalized prostate cell line RWPE and the prostate cancer cell lines LNCaP, Du145 NCI-H660 and PC3 were obtained from the American Type Culture Collection. Primary benign prostatic epithelial cells (PrEC) were obtained from Cambrex Bio Science. The prostate cancer cell lines C4-2B, LAPC4 and MDA-PCa 2B were provided by E. Keller. The prostate cancer cell line 22-RV1 was provided by J. Macoska. VCaP was derived from a vertebral metastasis from a patient with hormone-refractory metastatic prostate cancer[30].

For androgen stimulation experiments, LNCaP cells were grown in charcoal-stripped serum containing media for 24 h, before treatment for 24 h with 1% ethanol or 1 nM of methyltrienolone (R1881, NEN Life Science Products) dissolved in ethanol. For all samples, total RNA was isolated with Trizol (Invitrogen) according to the manufacturer's instructions.

Quantitative PCR. Q-PCR was performed using Power SYBR Green Mastermix (Applied Biosystems) on an Applied Biosystems 7300 Real Time PCR system as described[1,2]. All oligonucleotide primers were synthesized by Integrated DNA Technologies and are listed in Supplementary Table 2. HMBS and GAPDH[31], and PSA[32] primer sequences were as described. Androgen stimulation reactions were performed in quadruplicate, siRNA knockdown reactions were performed in triplicate and all other reactions were performed in duplicate.

RNA-ligase-mediated rapid amplification of cDNA ends. RLM-RACE was performed using the GeneRacer RLM-RACE kit (Invitrogen), according to the manufacturer's instructions as described[1,2]. To obtain the 5′ end of *ETV1*, first-strand cDNA was amplified with Platinum Taq High Fidelity (Invitrogen) using the GeneRacer 5′ primer and ETV1_exon4-5-r. For amplification from MET23, ETV1_exon7-r was used with the GeneRacer 5′ primer. Products were cloned and sequenced bidirectionally as described[1,2]. RLM-RACE cDNA was not used for other assays.

Fluorescence *in situ* hybridization. Interphase FISH on formalin-fixed paraffin-embedded tissue sections was performed as described[2]. A minimum of 50 nuclei per assay were evaluated. For metaphase FISH, spreads of LNCaP and MDA-PCa 2B were prepared using standard cytogenetic techniques. Slides were pre-treated in 2 × SSC for 2 min, 70% ethanol for 2 min and 100% ethanol for 2 min, and then air-dried. Slides and probes were co-denatured at 75 °C for 2 min, and hybridized overnight at 37 °C. Post-hybridization was in 0.5 × SSC at 42 °C for 5 min, followed by 3 washes in 1 × phosphate buffered saline with 0.1% Tween-20 (PBST). Fluorescent detection was performed using anti-digoxigenin conjugated to fluorescein (Roche Applied Science) and streptavidin conjugated to Alexa Fluor 594 (Invitrogen). Slides were counterstained and mounted in ProLong Gold Antifade Reagent with DAPI (Invitrogen). Slides were examined using a Zeiss Axio Imager Z1 fluorescence microscope (Zeiss) and imaged with a CCD camera using ISIS software (Metasystems). At least five metaphases were assessed, and reported aberrations were observed in all interpretable spreads. BACs (listed in Supplementary Table 3) were obtained from the BACPAC Resource Center, and probes were prepared as described[2]. Pre-labelled chromosome 7 centromere and 7p telomeric probes were obtained from Vysis. The integrity and correct localization of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes.

Tissue-specific expression. To determine the tissue-specific expression of 5′ fusion partners and genes at 14q13–q21, we interrogated the International Genomics Consortium's expO data set (http://expo.intgen.org/expo/public/downloaddata.jsp), consisting of expression profiles from 630 tumours of 29 distinct types, using the Oncomine database (http://www.oncomine.org)[27]. To interrogate the expression of *HERV-K_22q11.23*, which is not monitored by commercial array platforms, we queried the Lynx Therapeutics normal tissue MPSS data set (GSE1747) with the MPSS tag 'GATCTTTGTGACCTACT', which unambiguously identifies *HERV-K_22q11.23*, as described[33]. Descriptions of tumour types from the expO data set and the normal tissue types from the MPSS data set are provided in Supplementary Table 4.

Expression profiling. Expression profiling of LNCaP, C4-2B, RWPE-*ETV1* and RWPE-*GUS* cells was performed using the Agilent Whole Human Genome Oligo Microarray. Total RNA isolated using Trizol was purified using the Qiagen RNAeasy Micro kit. One microgram of total RNA was converted to cRNA and labelled according to the manufacturer's protocol (Agilent). Hybridizations were performed for 16 h at 65 °C, and arrays were scanned using an Agilent DNA microarray scanner. Images were analysed and data were extracted using Agilent Feature Extraction Software 9.1.3.1, with linear and lowess normalization performed for each array. For the LNCaP and C4-2B hybridizations, the reference for each cell line was pooled benign prostate total RNA (Clontech). A dye flip for each cell line was also performed. Features were ranked by average expression (log ratio) in the two LNCaP arrays divided by the average expression in the two C4-2B arrays after correction for the dye flip. For RWPE cells, four hybridizations were performed (duplicate RWPE-*ETV1*/RWPE-*GUS* and RWPE-*GUS*/RWPE-*ETV1* hybridizations). Over- and under-expressed signatures were generated by filtering to include only features with significant differential expression (P-value logratio < 0.01) in all four hybridizations and twofold average over- or under-expression (log ratio) after correction for the dye flip. Over- and under-expressed RWPE-*ETV1*/RWPE-*GUS* signatures were loaded into the Molecular Concepts Map[22], resulting in concepts containing 527 and 558 unique genes, respectively. Each signature was tested against all contained concepts in the Molecular Concepts Map for association using Fisher's exact test as described[10,22].

Southern hybridization. Genomic DNA (10 μg) from LNCaP, VCaP, pooled normal human male DNA (Promega) and normal placental DNA (Promega) was digested with *Eco*RI or *Pst*I (New England Biologicals) overnight. Fragments were resolved on a 0.8% agarose gel at 40 V overnight, transferred to Hybond NX nylon membrane, pre-hybridized, hybridized with probe and washed according to standard protocols. A series of 22 probes spanning the region of chromosome 7 implicated by FISH (between RP11-313C20 and RP11-703A4) were generated by PCR amplification with Platinum Taq High Fidelity on pooled normal human male genomic DNA (Supplementary Table 2 and Supplementary Fig. 5). Twenty-five nanograms of each probe was labelled with dCTP-^{32}P and used for hybridization.

Inverse PCR. To identify the *ETV1* breakpoint in LNCaP cells, we used an inverse PCR strategy based on the rearrangement identified by Southern blotting (probe A, Supplementary Table 2) as described previously[34] and shown in Supplementary Fig. 6. Primers A1, A2 and A3, which are reverse complemented from the wild-type sequence and are divergent to primers B1, B2 and B3, were used for inverse PCR on *Pst*I-digested and religated (to promote intramolecular ligation) LNCaP genomic DNA template. Nested PCRs were performed in the following order of primer combinations: A1–B1, A2–B2 and A3–B3. The Expand 20 kbplus PCR system (Roche) was used for amplifying the fusion product according to the manufacturer's suggestions. The enriched 3-kb band observed in nested PCRs was cloned into pCR8/GW/TOPO (Invitrogen), miniprep DNA was screened for inserts, and positive clones were sequenced (University of Michigan DNA Sequencing Core). The *ETV1* insertion was confirmed by PCR with Platinum Taq High Fidelity using primers from chromosomes 7 and 14 (Supplementary Table 2).

***In vitro* overexpression of *ETV1*.** cDNA of *ETV1*, as present in the *TMPRSS2–ETV1* fusion to the reported stop codon of *ETV1* (269–1521, NM_004956.3), was amplified by RT–PCR from MET26 (ref. 1) and TOPO cloned into the Gateway entry vector pCR8/GW/TOPO (Invitrogen), yielding pCR8-*ETV1*. To generate adenoviral and lentiviral constructs, pCR8-*ETV1* and a control entry clone (pENTR-*GUS*) were recombined with pAD/CMV/V5 (Invitrogen) and pLenti6/CMV/V5 (Invitrogen), respectively, using LR Clonase II (Invitrogen). Control pAD/CMV/*LACZ* clones were obtained from Invitrogen. Adenoviruses and lentiviruses were generated by the University of Michigan Vector Core. The benign immortalized prostate cell line RWPE was infected with lentiviruses expressing *ETV1* or *GUS*, and stable clones were generated by selection with blasticidin (Inivtrogen). Benign PrEC were infected with adenoviruses expressing *ETV1* or *LACZ*, because stable lines could not be generated in primary PrEC cells. Cell counts were estimated by treating cells with trypsin, and analysis was performed by a Coulter counter at the indicated time points in triplicate. For invasion assays, PrEC-*ETV1* and PrEC-*LACZ* (48 h after infection) or stable RWPE-*ETV1* and RWPE-*GUS* cells were used. Representative results from three separate experiments are shown in Fig. 3a and Supplementary Fig. 14.

***ETV1* knockdown.** For siRNA knockdown of *ETV1* in LNCaP cells, the individual siRNAs composing the Dharmacon SMARTpool against *ETV1* (MU-003801-01) were tested for *ETV1* knockdown by Q-PCR, and the most effective single siRNA (D-003801-05) was used for further experiments. siCONTROL non-targeting siRNA number 1 (D-001210-01) or siRNA against ETV1 was transfected into LNCaP cells using Oligofectamine (Invitrogen). After 24 h we performed a second identical transfection, and cells were harvested 24 h later for RNA isolation and invasion assays as described below. For shRNA knockdown of *ETV1* in LNCaP cells, the microRNA-adapted shRNA construct against *ETV1* from the pMS2 retroviral vector (V2HS_61929, Open Biosystems) was cloned into an empty pGIPZ lentiviral vector (RHS4349, Open Biosystems) according to the manufacturer's protocol. pGIPZ lentiviruses with microRNA-adapted shRNAs against *ETV1* or a non-silencing control (RHS4346) were generated by the University of Michigan Vector Core. LNCaP cells were infected with lentiviruses, and 48 h later cells were used for invasion assays as described below. Representative results from six independent experiments are reported.

Invasion assays. Equal numbers of the indicated cells were seeded onto the basement membrane matrix (EC matrix, Chemicon) present in the insert of a 24-well culture plate, with fetal bovine serum added to the lower chamber as a chemoattractant. After 48 h, non-invading cells and the EC matrix were removed by a cotton swab. Invaded cells were stained with crystal violet and photographed. The inserts were treated with 10% acetic acid and absorbance was measured at 560 nm.

FACS cell cycle analysis. RWPE-*ETV1* and RWPE-*GUS* cells were assessed by FACS for cell-cycle characterization. Cells were washed with 2 × PBS, and approximately 2×10^6 cells were resuspended in PBS before fixation in 70% ethanol. Pelleted cells were washed and treated with RNase (100 μg ml^{-1} final concentration) and propidium iodide (10 μg ml^{-1} final concentration) at 37 °C for 30 min. Stained cells were analysed on a LSR II flow cytometer (BD Biosciences) running FACSDivia, and cell-cycle phases were calculated using ModFit LT (Verity Software House).

Soft agar assay. A 0.6% (w/v) bottom layer of low-melting-point agarose in normal medium was prepared in six-well culture plates. On top, a layer of 0.3% agarose containing 1×10^4 RWPE-*GUS*, RWPE-*ETV1* or *DU145* (positive control) cells was placed. After 12 d, foci were stained with crystal violet and were counted.

Transgenic *ETV1* mice. For *in vivo* overexpression of *ETV1*, a carboxy-terminal 3×Flag-epitope-tagged construct was generated by PCR using pCR8-*ETV1* as the template, with the reverse primer encoding a triple Flag tag before the stop codon. The product was TOPO cloned into pCR8. To generate a prostate-specific *ETV1* transgenic construct, 3×Flag-*ETV1* was inserted into pBSII (Stratagene) downstream of a modified small composite probasin promoter (ARR2Pb) and upstream of a bovine growth hormone polyA site (PA-BGH). The ARR2Pb sequence contains the original probasin sequence Pb (426/+28) plus two additional androgen response elements[23]. The construct was sequenced and androgen was used to test for promoter inducibility in LNCaP cells on transient transfection before microinjection into FVB mouse eggs. The ARR2Pb-*ETV1* plasmid was linearized with *Pvu*I/*Kpn*I//*Sac*II and microinjected into fertilized FVB mouse eggs, and was then surgically transplanted into a pseudo-pregnant female by the University of Michigan Transgenic Animal Model Core. Transgenic founders were screened by PCR using genomic DNA isolated from tail snips. Multiple ARR2Pb-*ETV1* transgenic founders were obtained and crossed with FVB mice, and transgene-positive male mice offspring were killed at various time points.

Prostates from transgenic mice were dissected using a Nikon dissection scope, fixed in 10% buffered formalin and embedded in paraffin. Five-micrometre sections were stained with haematoxylin and eosin, and evaluated by three pathologists (R.M., M.A.R. and R.B.S.) according to the criteria provided in the Consensus Report from the Bar Harbor Meeting of the Mouse Models of Human Cancer Consortium Prostate Pathology Committee[24].

29. Rubin, M. A. *et al.* Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. *Clin. Cancer Res.* **6**, 1038–1045 (2000).
30. Korenchuk, S. *et al.* VCaP, a cell-based model system of human prostate cancer. *In Vivo* **15**, 163–168 (2001).
31. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).
32. Specht, K. *et al.* Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am. J. Pathol.* **158**, 419–429 (2001).
33. Stauffer, Y., Theiler, G., Sperisen, P., Lebedev, Y. & Jongeneel, C. V. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun.* **4**, 2 (2004).
34. Wiemels, J. L. & Greaves, M. Structure and possible mechanisms of *TEL–AML1* gene fusions in childhood acute lymphoblastic leukemia. *Cancer Res.* **59**, 4075–4082 (1999).

LETTERS

Behavioural improvements with thalamic stimulation after severe traumatic brain injury

N. D. Schiff[1], J. T. Giacino[2,3], K. Kalmar[2], J. D. Victor[1], K. Baker[4], M. Gerber[2], B. Fritz[2], B. Eisenberg[2], J. O'Connor[2], E. J. Kobylarz[1], S. Farris[4], A. Machado[4], C. McCagg[2], F. Plum[1], J. J. Fins[5] & A. R. Rezai[4]

Widespread loss of cerebral connectivity is assumed to underlie the failure of brain mechanisms that support communication and goal-directed behaviour following severe traumatic brain injury. Disorders of consciousness that persist for longer than 12 months after severe traumatic brain injury are generally considered to be immutable; no treatment has been shown to accelerate recovery or improve functional outcome in such cases[1,2]. Recent studies have shown unexpected preservation of large-scale cerebral networks in patients in the minimally conscious state (MCS)[3,4], a condition that is characterized by intermittent evidence of awareness of self or the environment[5]. These findings indicate that there might be residual functional capacity in some patients that could be supported by therapeutic interventions. We hypothesize that further recovery in some patients in the MCS is limited by chronic under-activation of potentially recruitable large-scale networks. Here, in a 6-month double-blind alternating crossover study, we show that bilateral deep brain electrical stimulation (DBS) of the central thalamus modulates behavioural responsiveness in a patient who remained in MCS for 6 yr following traumatic brain injury before the intervention. The frequency of specific cognitively mediated behaviours (primary outcome measures) and functional limb control and oral feeding (secondary outcome measures) increased during periods in which DBS was on as compared with periods in which it was off. Logistic regression modelling shows a statistical linkage between the observed functional improvements and recent stimulation history. We interpret the DBS effects as compensating for a loss of arousal regulation that is normally controlled by the frontal lobe in the intact brain. These findings provide evidence that DBS can promote significant late functional recovery from severe traumatic brain injury. Our observations, years after the injury occurred, challenge the existing practice of early treatment discontinuation for patients with only inconsistent interactive behaviours and motivate further research to develop therapeutic interventions.

Severe traumatic brain injury typically results in *en passant* injuries to thalamic and midbrain structures that are essential parts of the forebrain arousal regulation system[6–11]. We sought to determine whether DBS in the central thalamus could promote behavioural responsiveness in a patient in a chronic MCS by approximating the normal role of mesial frontal cortical and brainstem inputs, which adjust firing rates in central thalamic neurons to regulate cognitive effort and maintain brain metabolic activity during normal wakefulness[10,11].

As part of a multi-institutional, FDA- and IRB-approved clinical trial, we implanted DBS electrodes bilaterally within the central thalamus of a 38-yr-old male who remained in an MCS following a severe traumatic brain injury (see Supplementary Information). Over a two-year course of inpatient rehabilitation and four subsequent years in a nursing home, he failed to recover consistent command-following or communication ability and remained non-verbal. Six-and-a-half years after the injury, the patient was re-admitted to an inpatient rehabilitation unit for comprehensive re-evaluation and rehabilitation. Although he remained unable to communicate reliably, functional MRI showed preservation of a large-scale, bi-hemispheric cerebral language network, indicating that a substrate for further recovery might exist[4]. Additional studies using positron emission tomography showed that the patient's resting global cerebral metabolism was markedly reduced. These observations supported our hypothesis that the patient's inconsistent behavioural responsiveness and communication reflected a global reduction in neuronal activity resulting from widespread de-afferentation and compression injuries to the thalamus and midbrain[4].

We used a single-subject, multiple baseline design to investigate the effects of DBS using a priori statistical evaluation of preselected behavioural metrics. A presurgical baseline established the patient's level of responsiveness before surgery. Post-surgical assessments were conducted within 48 h and during a 2-month period preceding a DBS titration phase in which the patient was exposed to varying patterns of stimulation, to allow us to identify optimal behavioural responses. After the titration phase, a six-month double-blinded crossover phase began, in which DBS was alternated between being turned on and turned off every 30 days (Fig. 1). A multidisciplinary neuro-rehabilitation team performed all evaluations using standardized assessment procedures.

To assess the effects of DBS, we prospectively chose the JFK Coma Recovery Scale — Revised (CRS-R), a measure of neurobehavioural function that has been validated in patients with disorders of consciousness[12,13] (Supplementary Fig. 1). We also developed three secondary outcome measures that assessed object naming, purposeful upper extremity limb movement and oral feeding to characterize behavioural changes more fully (see Supplementary Information). A comprehensive inpatient rehabilitation program was initiated four months before surgery and continued without modification throughout the study (Figs 1, 2a,b). This program consisted of physical, occupational, speech and recreational therapies and did not differ from the patient's initial course of rehabilitation, which had been completed four years earlier.

CRS-R evaluations conducted over a three-week presurgical baseline verified that the patient's neurobehavioural status was stable. Three subscales of the CRS-R were subsequently selected as the primary outcome measures. Scores on the Arousal subscale indicated that the patient could not consistently respond to basic verbal commands.

[1]Department of Neurology & Neuroscience, Weill Cornell Medical College, New York, New York 10021, USA. [2]JFK Johnson Rehabilitation Institute, Edison, New Jersey 08818, USA. [3]New Jersey Neuroscience Institute, Edison, New Jersey 08818, USA. [4]Center for Neurologic Restoration, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA. [5]Division of Medical Ethics, Weill Cornell Medical College, New York, New York 10021, USA.

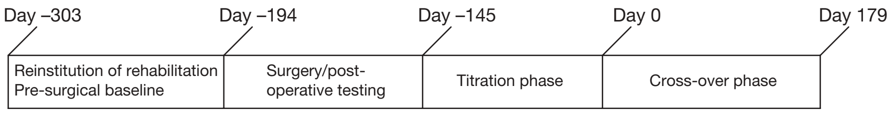


Figure 1 | Study timeline. Timeline illustrating the 483 days of data collection across the different phases of the study.

Motor and Communication scores demonstrated inability to use familiar objects in a purposeful manner or to respond reliably to simple yes/no questions, respectively (Supplementary Table 1).

We then implanted the DBS electrodes, targeting the anterior intralaminar thalamic nuclei and adjacent paralaminar regions of thalamic association nuclei bilaterally (see Supplementary Information). Both electrodes were positioned with each of the four individual contacts within the central lateral nucleus, paralaminar regions of the median dorsalis, and the posterior-medial aspect of the centromedian/parafascicularis nucleus complex. This placement maximized coverage of thalamic regions with strong calbindin protein staining[14]. These calbindin-positive neurons project to supragranular cortical regions[15], consistent with the idea that they have a parallel role in cerebral activation that is comparable to that of projections from the brainstem arousal systems[16].

We evaluated electrical stimulation at each electrode contact within 48 h of surgery to identify potential adverse effects and to determine voltage thresholds for behavioural changes. Stimulation produced acute changes in arousal including increased heart rate, well-sustained eye opening and rapid bilateral head-turning to a voice[17]. Such effects were observed for all electrode contacts tested, and each electrode had a preferred contact with the lowest voltage threshold for eliciting a response. We further determined a specific configuration of electrode cathode/anode geometries (monopolar versus bipolar) for each electrode that minimized this voltage threshold. We assessed behaviours over 20–30-min periods of bilateral DBS that alternated between on and off. During object naming, speech remained unintelligible and was limited to episodes of incomprehensible word-mouthing. The frequency of limb movements that involved social gesture and object use was significantly higher with DBS on than with it off (Supplementary Table 2), although we never observed fully executed movement sequences. On postoperative day 2, we recorded cortical potentials evoked by electrical stimuli delivered to the individual contacts of each DBS lead using time-locked averaging of the scalp EEG (Supplementary Fig. 7). The total exposure to DBS during this phase was approximately 2 h, and there was no further exposure to DBS for the next 2 months (Fig. 2b).

The titration phase began 50 days after surgery (Figs 1, 2b) and continued for 18 weeks. The CRS-R, object naming and limb movement protocols were administered weekly to test a range of different stimulation frequencies (70–250 Hz) and intensities (0–5 V). The duration of alternating time on and off stimulation was gradually increased from three to fourteen days. The frequency range explored was guided by ongoing empirical observations of the patient and previous results in primate and rodent experiments that showed behavioural facilitation by stimulation of the central thalamus using frequencies of 50–100 Hz[18,19].

We noted several qualitative changes in behaviour during DBS titration that correlated with abrupt changes in CRS-R subscale scores (Fig. 2a). These changes were observed shortly after the onset of continuous bilateral stimulation on day –145 (Fig. 2b). Longer periods of eye opening and increased responsiveness to command were reflected in increased CRS-R Arousal scores. On day –143, the patient showed the first instances of functional object use on the Motor subscale (score = 6) and intelligible verbalization on the Oromotor subscale (score = 3). No previous episodes of intelligible verbalization had been observed during a series of 33 evaluations conducted across the first 6 months of observation. These behavioural improvements (and other improvements, see Supplementary Information) temporally coincided with the onset of DBS (Fig. 2a) and did not emerge until 160 days after initiation of the rehabilitation program, indicating that they were primarily attributable to the DBS.

After titration testing, we chose a final set of stimulation parameters for each electrode (100 Hz, right side–bipolar field, 4 V; left side–monopolar field, 4 V), reflecting considerations of behavioural observations and of battery life. The patient then began the crossover phase. Data were collected daily by the same team that performed the baseline assessments. Comparison of the DBS on and off conditions showed that the patient received the maximal score on the Arousal subscale significantly more often during the DBS-on periods (Fig. 3; $P < 0.001$, Pearson Chi-square (two-tail), Systat). There was no significant difference in performance on the CRS-R Motor and Communication subscale scores throughout the crossover phase, probably reflecting a ceiling effect on both subscales. The patient's ability to execute complete functional limb movement sequences (for example, bringing a cup to his mouth) also significantly improved with DBS on ($P < 0.001$, Pearson Chi-square (two-tail), Systat), as did his ability independently to chew and swallow a bolus of food placed on his tongue (Fig. 3; $P < 0.001$, Pearson Chi-square (two-tail), Systat). There was no significant difference in object-naming ability.

To distinguish the effects of DBS from those produced by the rehabilitation program or by repeated exposure to the testing procedures, we performed a logistic regression analysis (see Supplementary Information). During the crossover phase, traditional rehabilitation efforts continued, but exposure to DBS alternated on a monthly basis. Thus, behavioural changes that were temporally linked to DBS should be modulated in an approximately square-wave fashion, whereas effects due to rehabilitation or the evaluation process itself are expected to grow gradually over time. The regression model has three parameters (*A*, *B* and *C*), each determining the strength of a potentially contributing influence. The *A*-term represents the overall probability of obtaining the higher rating on a given scale. (For this analysis, rating scales were dichotomized into two classes of approximately equal size.) The *B*-term allows this probability to increase or decrease gradually as a function of elapsed time. The *C*-term allows

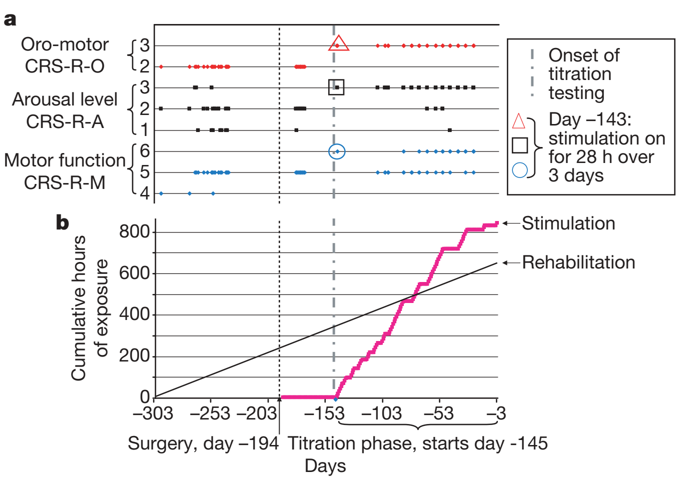


Figure 2 | Qualitative changes in behaviour on CRS-R subscales associated with titration phase and cumulative hours of rehabilitation and brain stimulation. a, The CRS-R subscale scores are shown from enrolment (Day –309) to the end of the titration phase (Day –1). Abrupt changes in CRS-R subscale scores were seen after the onset of the titration phase (Day –145). **b**, Cumulative hours of rehabilitation and electrical brain stimulation are shown across the same time period. Exposure to rehabilitation hours is constant at 3 h per day, 5 days per week (estimated as 660 h over 44 weeks, black line). The accumulation of hours of continuous stimulation is shown in magenta.

this probability to be modulated by the recent stimulation history. The model was applied, in separate analyses, to the three behavioural measures that improved significantly during the crossover phase (CRS-R Arousal, limb control and oral feeding; Fig. 4, Supplementary Fig. 2a,b and Supplementary Table 3). For all analyses, including the *C*-term (reflecting behavioural modulation tracking the stimulation course) in the model leads to a significant improvement ($P < 0.005$ or $P < 0.001$; bootstrap confidence interval on *C*-term significant at $P < 0.01$ for all models). Inclusion of the *B*-term, which reflects gradual change, produces only negligible improvement in the model (not significant for arousal and oral feeding, $P < 0.05$ for limb control but 95% confidence interval for *B*-term includes zero). These analyses indicate that nearly all of the systematic changes in ratings were temporally linked to DBS stimulation, and could not be accounted for by a gradual improvement over time. In addition, logistic regression analyses showed that improvements in intelligible verbalization and limb control were also directly linked to the DBS-on condition during the titration phase (Supplementary Information).

The observed improvements in arousal level, motor control and behavioural persistence might reflect direct activation of frontal cortical and basal ganglia systems that are innervated by neurons within the anterior intralaminar regions and adjacent paralaminar regions of thalamic association nuclei. These neurons act as a key intermediary system and common final pathway for brainstem arousal systems and frontal lobe regions that exert executive control of effort regulation, working memory, selective attention and vigilance[6–11].

Anatomical studies in nonhuman primates have shown that the central lateral nucleus and surrounding paralaminar regions widely innervate the premotor and supplementary motor cortices, providing several monosynaptic pathways that might have been stimulated in our study[20]. We conclude that mesial frontal regions within the supplementary motor area, anterior cingulate cortex and other frontal cortical regions were activated by the electrical stimulation. The frontocentral predominance of cortical activations that were associated with activation of the left DBS lead contacts support this localization, but we cannot exclude the possibility of contributions from surrounding thalamic neurons (see Supplementary Information).

Our focus on patients in a chronic MCS with widely preserved brain structure and clear evidence of interactive behaviour (command following, verbalization and inconsistent communication) is motivated by two important considerations. First, we propose the use of central thalamic DBS to improve the arousal regulation of functionally connected but inconsistently active cerebral networks that might be present in some patients in an MCS but absent in patients in a permanent vegetative state[21]. Second, for patients in an MCS who have not yet recovered reliable communication or functional movements, improvements in response initiation and persistence might restore these abilities.

The behavioural improvements described in this report are notable, given their late emergence and potential functional significance. However, the generalizability of the results is unknown, and expectations raised by this report should be tempered. In particular, our patient followed commands and showed intact language networks in neuroimaging studies[4]. These characteristics will not be shared by all patients in an MCS. Moreover, unknown aspects of this particular subject's brain injury might have influenced his response to DBS. Nonetheless, replication of these findings could have important implications for clinical practice. Although some patients in MCS show clear verbal or gestural 'yes/no' responses, these do not occur consistently enough to be considered reliable. Our patient's recovered ability to interact consistently and meaningfully with others was cited by members of his family as the most important change observed. The restoration of communication also allowed the patient to assume a more active (and interactive) role in his treatment.

These considerations motivate the development of a neuropalliative ethic to establish proportionate goals of care to help

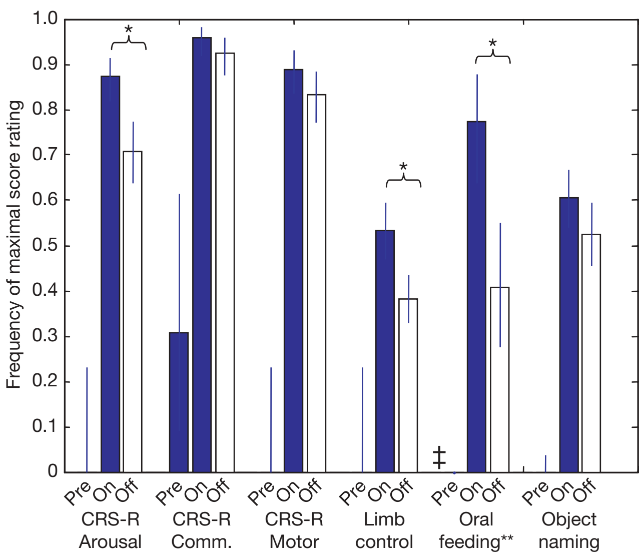


Figure 3 | Comparison of pre-surgical baselines and DBS on and DBS off periods during the crossover phase. Pre-surgical baselines (14 consecutive observations) and crossover phase observations are displayed with 95% confidence intervals for binomial distributions with *n* observations for three CRS-R subscales (Arousal, $n = 185$ off, 189 on; Communication (Comm.), $n = 185$ off, 189 on; Motor, $n = 185$ off, 189 on) and secondary measures (limb movement, $n = 336$ off, 261 on; oral feeding, $n = 54$ off, 53 on; object naming, $n = 206$ off, 235 on). Asterisks indicate significant differences between on versus off for CRS-R Arousal, limb movement and oral feeding ($P < 0.001$ established by Pearson Chi-square (two-tail)). Daggers indicate that oral feeding data were not available before titration; double asterisk indicates that scores 1 and 2 are combined for dichotomy. Pre-surgical baselines for limb control and object naming were converted from CRS-R Motor and CRS-R Oro-motor scores, which remained below the maximal rating across the 14 consecutive observations in the pre-surgical baseline.

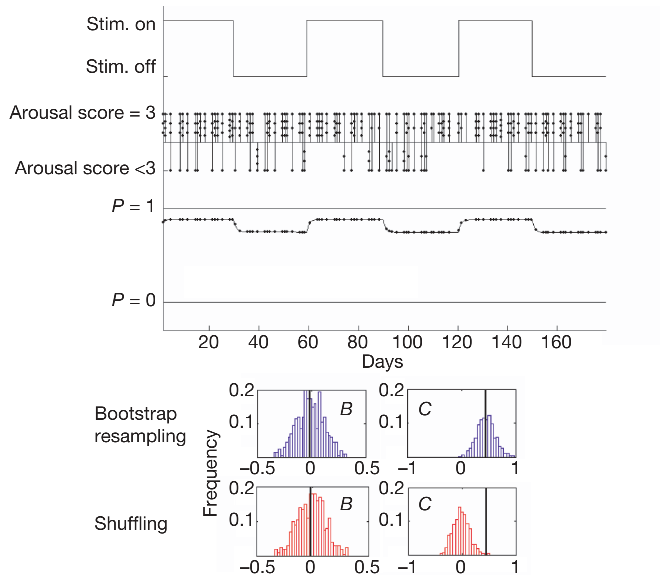


Figure 4 | Logistic regression model of crossover phase arousal data. Figure shows stimulation and behavioural time series, and best-fitting probability model for the behavioural data selected by logistic regression analyses. Single dots on vertical lines within behavioural time series indicate multiple observations on the same day, dots on probability model indicate times of observations. Histograms show bootstrap (blue) and shuffled (red) estimates (500 each) of the coefficients for a linear trend (*B*) and the stimulation history (*C*). Data-derived estimates of the coefficients are shown as black lines. $P = 1$ indicates CRS-R arousal score of 3, $P = 0$ indicates other response categories.

families to balance the potential for improvement against associated burdens while being guided by the patient's previously articulated preferences[22]. Confirmation of these findings in other patients might influence the current practice of excluding individuals with inconsistent behavioural responsiveness from structured rehabilitation programs. Our findings should motivate research to elucidate the mechanisms of recovery and to facilitate the identification of patients who might benefit from neuromodulatory interventions[23].

METHODS SUMMARY

A patient meeting the diagnostic criteria for an MCS proposed by Giacino *et al.*[5] was enrolled as part of a clinical trial of central thalamic DBS. After determination of eligibility, the patient entered into a four-month phase of behavioural evaluation with formal behavioural assessments using the JFK CRS-R[12] to identify subscales that failed to show a ceiling effect. Three subscales (Arousal, Communication and Motor) were prospectively chosen as primary outcome variables in view of the inconsistency in performance on these measures. Subsequently, the patient underwent implantation of bilateral DBS electrodes and post-operative physiological evaluations, including the recording of evoked potentials generated by the stimulation of individual electrode contacts before internalization of electrodes and connection to pulse generators. A DBS titration phase began 50 days after implantation, during which different combinations of frequency, intensity, electrode contact activation and periods of on and off times were tested. At the end of the titration phase, a six-month double-blinded alternating on/off crossover phase began, using optimal parameters that were selected during the titration phase. Secondary outcome measures were developed during the titration phase and included object naming, limb control and oral feeding indices. All primary and secondary outcome measures were tracked prospectively during the crossover phase. During the entire study (from enrolment to completion of the crossover phase), the patient continued to undergo all routine rehabilitation activities. For analysis of DBS data obtained during the on/off crossover phase, we compared binomial data using the Pearson Chi-square (two-tail) test, or the Fisher Exact test when the sample size was insufficient to use the former (Systat). We obtained confidence intervals for binomials using 'binofit' in Matlab 7.0 (Mathworks). Small sample comparisons were completed using the Wilcoxon ranksum test (Matlab 7.0). Logistic regression models of the crossover data were fit using the 'fmincon' routine in Matlab 7.0.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 April; accepted 22 June 2007.

1. Lombardi, F., Taricco, M., DeTanti, A., Telaro, E. & Liberati, A. Sensory stimulation for brain injured individuals in coma or vegetative state. *Cochrane Database Syst. Rev.* 2, CD001427 (2002).
2. Giacino, J. & Whyte, J. The vegetative state and minimally conscious state: current knowledge and remaining questions. *J. Head Trauma Rehabil.* 20, 30–50 (2005).
3. Boly, M. *et al.* Auditory processing in severely brain injured patients: differences between the minimally conscious state and the persistent vegetative state. *Arch. Neurol.* 61, 233–238 (2004).
4. Schiff, N. D. *et al.* fMRI reveals large-scale network activation in minimally conscious patients. *Neurology* 64, 514–523 (2005).
5. Giacino, J. T. *et al.* The minimally conscious state: definition and diagnostic criteria. *Neurology* 58, 349–353 (2002).
6. Schiff, N. D. & Purpura, K. P. Towards a neurophysiological basis for cognitive neuromodulation. *Thalamus Relat. Syst.* 2, 55–69 (2002).
7. Steriade, M. & Glenn, L. L. Neocortical and caudate projections of intralaminar thalamic neurons and their synaptic excitation from midbrain reticular core. *J. Neurophysiol.* 48, 352–371 (1982).
8. van der Werf, Y. D., Witter, M. P. & Groenewegen, H. J. The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Res. Brain Res. Rev.* 39, 107–140 (2002).
9. Wyder, M. T., Massoglia, D. P. & Stanford, T. R. Contextual modulation of central thalamic delay-period activity: representation of visual and saccadic goals. *J. Neurophysiol.* 91, 2628–2648 (2004).
10. Kinomura, S., Larssen, J., Gulyas, B. & Roland, P. E. Activation by attention of the human reticular formation and thalamic intralaminar nuclei. *Science* 271, 512–515 (1996).
11. Paus, T. *et al.* Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. *J. Cogn. Neurosci.* 9, 392–408 (1997).
12. Giacino, J. T., Kalmar, K. & Whyte, J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch. Phys. Med. Rehabil.* 85, 2020–2029 (2004).
13. Giacino, J. T., Kezmarsky, M. A., DeLuca, J. & Cicerone, K. D. Monitoring rate of recovery to predict outcome in minimally responsive patients. *Arch. Phys. Med. Rehabil.* 72, 897–901 (1991).
14. Munkle, M. C., Waldvogel, H. J. & Faull, R. L. The distribution of calbindin, calretinin and parvalbumin immunoreactivity in the human thalamus. *J. Chem. Neuroanat.* 19, 155–173 (2000).
15. Jones, E. G. The thalamic matrix and thalamocortical synchrony. *Trends Neurosci.* 24, 595–601 (2001).
16. Llinas, R. R., Leznik, E. & Urbano, F. J. Temporal binding via cortical coincidence detection of specific and nonspecific thalamocortical inputs: a voltage-dependent dye-imaging study in mouse brain slices. *Proc. Natl Acad. Sci. USA* 99, 449–454 (2002).
17. Pfaff, D. *Brain Arousal and Information Theory* (Harvard Univ. Press, Cambridge, Massachusetts, 2005).
18. Shirvalkar, P., Seth, M., Schiff, N. D. & Herrera, D. G. Cognitive enhancement through central thalamic deep brain stimulation. *Proc. Natl Acad. Sci. USA* 103, 17007–17012 (2006).
19. Schiff, N. D., Hudson, A. E. & Purpura, K. P. Modeling wakeful unresponsiveness: characterization and microstimulation of the central thalamus. *Soc. Neurosci. 31st Annual Meeting* Abstr. 62.12 (2002).
20. Morel, A., Liu, J., Wannier, T., Jeanmonod, D. & Rouiller, E. M. Divergence and convergence of thalamocortical projections to premotor and supplementary motor cortex: a multiple tracing study in the macaque monkey. *Eur. J. Neurosci.* 21, 1007–1029 (2005).
21. Jennett, B. *et al.* Neuropathology in vegetative and severely disabled patients after head injury. *Neurology* 56, 486–490 (2001).
22. Fins, J. J. Constructing an ethical stereotaxy for severe brain injury: balancing risks, benefits and access. *Nature Rev. Neurosci.* 4, 323–327 (2003).
23. Fins, J. J. A proposed ethical framework for interventional cognitive neuroscience: a consideration of deep brain stimulation in impaired consciousness. *Neurol. Res.* 22, 273–278 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for this study came from the NINDS, Charles A. Dana Foundation, Cleveland Clinic Foundation Brain Neuromodulation Center, Ohio Department of Development BRTT and Third Frontier Program, Jane and Lee Seidman Neuromodulation Research Fund, Cleveland Clinic Innovations, IntElect Medical and the NIDDR. We thank L. Turkstra, E. Montgomery, A. Sharan, D. Vegh and K. Purpura for contributions in planning stages of this work; E. Bagiella for statistical consultation; and C. Allen, D. Pfaff, J. Whyte, M. Gizzi and R. Grewal for expert editorial review.

Author Contributions N.D.S., J.T.G. and A.R.R. contributed equally to the design, implementation and scholarly preparation of this work. N.D.S. and J.T.G. wrote the manuscript. N.D.S., J.T.G. and A.R.R. acted as principal investigators at their performance sites and participated in all phases of the study including evaluation of data and preparation of all parts of the manuscript. A.R.R. and N.D.S. acted as co-principal investigators for the Investigational Device Exemption (Food and Drug Administration, IDE) covering the use of the brain stimulation methods. N.D.S. acted as principal investigator for the initiation of the project and the development grant that formed the basis of the studies. J.T.G. developed, organized and supervised the collection and primary analysis of behavioural data along with K.K. and an independent statistical consultant (E. Bagiella, Columbia University). J.D.V. assisted with the development of the study design, developed the logistic regression models and supervised this data analysis, and assisted in the preparation of manuscript. K.B. analysed the evoked potential response data and prepared the manuscript presentation of the results; K.B. and N.D.S. collected the evoked potential data. M.G., B.F., B.E. and J.O. collected behavioural data and assisted in the development of secondary outcome measures. C.M. served as the patient's primary care physician and supervised and administered the deep brain electrical stimulation of the patient. Owing to role sequestration, she had no role in data collection or analysis. A.R.R. organized, supervised and carried out the neurosurgical planning, procedures and follow-up, supervised the programming of the neurostimulation and assisted in preparation of the manuscript. J.J.F. developed formal consent procedures and conceptual frameworks with the study primary investigators to address ethical considerations arising in all aspects of the study and preparation of the manuscript. E.J.K., A.M. and K.B. participated in the pre-surgical evaluation, operative and post-operative patient evaluations and neurophysiological evaluations. S.F. provided expert assistance in the design and evaluation of the stimulation protocols. F.P. acted as principal investigator for the planning study for the first year and as a senior advisor throughout the development of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D.S. (nds2001@mail.med.cornell.edu).

doi:10.1038/nature06041

METHODS

Overview. Pre-surgical assessments were conducted to enable pre-selection of the three primary outcome variables. After completion of the pre-surgical baseline, the patient underwent implantation of bilateral DBS electrodes. The electrodes remained off until initiation of a DBS titration period which began 50 days after implantation, except for a 2-day period of stimulation testing in the immediate post-operative period. Different combinations of frequency, intensity, electrode contact activation and periods of on and off times were tested (see below). At the end of the titration phase, a six-month double-blinded alternating on/off crossover trial began using optimal parameters selected during the titration phase. The Arousal, Motor and Communication subscales of the JFK Coma Recovery Scale — Revised served as the primary outcome measures. The secondary outcome measures were developed during the titration phase and included object naming, limb control and oral feeding indices. All measures were tracked prospectively during the crossover phase.

The subject. A 38-yr-old right-handed man who sustained a closed head injury after an assault 6 yr before the study was enrolled after fulfilling the entry criteria. The patient remained in an MCS according to the criteria proposed in ref. 5. The patient remained medically stable, and structural brain imaging showed no evidence of significant bilateral frontal lesions or injury to the left inferior frontal operculum or the left posterior temporal-parietal region. During the entire study, the patient continued to undergo all routine rehabilitation activities.

Pre-surgical evaluation phase. After selection into the study and determination of eligibility, the patient entered a four-month phase of behavioural evaluation. Towards the end of this phase, formal behavioural assessments using the CRS-R were carried out over a three-week period (twice daily) to identify subscales that failed to show a ceiling effect. Three subscales (Arousal, Communication and Motor) were prospectively chosen as primary outcome variables in view of the patient's inconsistency in performance on these measures.

Electrode implantation and stimulation. After selection for DBS surgery, the patient underwent pre-operative physiological evaluations, implantation of DBS electrodes and post-operative physiological evaluations including recordings from the implanted DBS electrodes before internalization of electrodes and connection to pulse generators.

Post-operative testing phase. Over a two-day period following implantation of the electrode leads, the individual contacts of each electrode were tested to identify voltage thresholds for behavioural effects. During this time, DBS-evoked responses were recorded.

Titration phase. During the titration phase, we obtained CRS-R scores and developed the secondary outcome measures while exploring different stimulation parameters. Observations included comparisons of the frequency of the highest CRS-R subscale scores obtained before surgery with those obtained during titration, and weekly assessments of object naming, functional movements and oral feeding performed by physical, speech, occupational and recreational therapists.

Crossover phase. After the pre-treatment titration phase, we chose a specific parameter set for stimulation for each electrode. The patient then entered into a double-blinded alternating crossover phase of DBS on and off using only these stimulation parameters.

Statistics. All small sample comparisons were completed using the Wilcoxon ranksum test (Matlab 7.0, Mathworks). We compared binomial data using the Pearson Chi-square (two-tail) test, or the Fisher Exact test when the sample size was insufficient to use the former (Systat). We obtained confidence intervals for binomial data using 'binofit' in Matlab 7.0. Logistic regression models were fit using the 'fmincon' routine in Matlab 7.0.

nature Vol 448|2 August 2007|doi:10.1038/nature06001

LETTERS

Common effector processing mediates cell-specific responses to stimuli

Kathryn Miller-Jensen[1,2]*, Kevin A. Janes[1,3]*, Joan S. Brugge[3] & Douglas A. Lauffenburger[1,2,4]

The fundamental components of many signalling pathways are common to all cells[1–3]. However, stimulating or perturbing the intracellular network often causes distinct phenotypes that are specific to a given cell type[4,5]. This 'cell specificity' presents a challenge in understanding how intracellular networks regulate cell behaviour and an obstacle to developing drugs that treat signalling dysfunctions[6,7]. Here we apply a systems-modelling approach[8] to investigate how cell-specific signalling events are integrated through effector proteins to cause cell-specific outcomes. We focus on the synergy between tumour necrosis factor and an adenoviral vector as a therapeutically relevant stimulus that induces cell-specific responses[9–11]. By constructing models that estimate how kinase-signalling events are processed into phenotypes through effector substrates, we find that accurate predictions of cell specificity are possible when different cell types share a common 'effector-processing' mechanism. Partial-least-squares regression models based on common effector processing accurately predict cell-specific apoptosis, chemokine release, gene induction, and drug sensitivity across divergent epithelial cell lines. We conclude that cell specificity originates from the differential activation of kinases and other upstream transducers, which together enable different cell types to use common effectors to generate diverse outcomes. The common processing of network signals by downstream effectors points towards an important cell biological principle, which can be applied to the understanding of cell-specific responses to targeted drug therapies[6].

Signalling networks arise from complex modular interactions among the adaptors, enzymes, and other transducer proteins that together transmit information[12] (Fig. 1a). Experimental[13] and literature-based[14] reconstructions of signalling networks often assume that there exists a common network whose connectivity is the same for all cells[15]. However, cellular responses to stimuli are usually specific to a given cell type. How are cell-specific responses achieved through a common signalling network? To answer this question, it is important to consider the downstream effector substrates (including cytoplasmic targets, transcription factors and terminal enzymes) that ultimately 'process' the activation of transducer proteins into cell-specific phenotypes (Fig. 1a).

We investigated how effector substrates process signalling-network information by using a defined experimental setup with clear, cell-specific outcomes. Adenoviral gene-therapy vectors (Adv) synergize with the cytokine tumour necrosis factor (TNF) in many cellular contexts[11,16,17]. The binding and entry of Adv activates many pathways that are shared with TNF (Supplementary Table 1). In epithelial cells, apoptosis caused by the synergistic combination of Adv and TNF occurs together with Adv-induced changes in activation of the TNF signalling network[11]. Interestingly, the extent

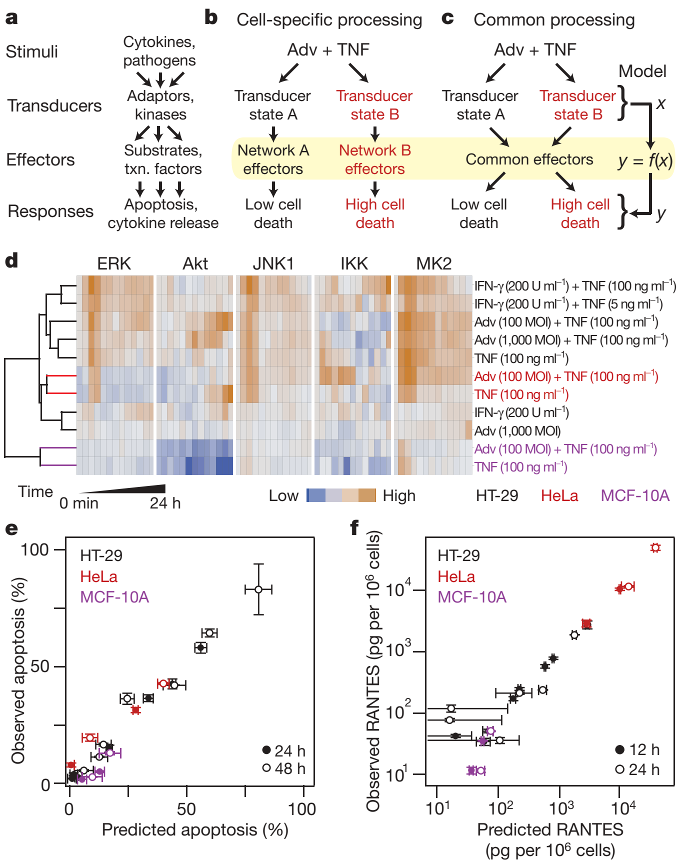


Figure 1 | Test of a common effector-processing hypothesis in epithelial cells treated with Adv plus TNF. a, Flow of information from stimuli through transducers and effectors to responses. Txn factors, transcription factors. **b, c,** Cell-specific (**b**) and common effector-processing mechanisms (**c**) linking treatment with Adv plus TNF to apoptosis. A model of effector processing, $f(x)$, links transducer activation to responses. **d,** Time-dependent kinase activities in HT-29 (black), HeLa (red) and MCF-10A (purple) cells treated with different combinations of Adv, IFN-γ and TNF[18,19]. Data are presented as the mean of duplicate (Adv) or triplicate (IFN-γ) biological samples (see Methods and Supplementary Table 5 for details). MOI, multiplicity of infection. **e, f,** Correlation between predictions of the common-processing model and measurements of apoptosis (**e**) or RANTES secretion (**f**). Model training with HT-29 cells (black) is compared to model predictions of HeLa (red) and MCF-10A (purple) cells at 24 h (filled) and 48 h (open) after treatment for apoptosis or 12 h (filled) and 24 h (open) after treatment for RANTES secretion. Experimental data are presented as the mean ± s.e. ($n = 3$), and model uncertainties are estimated by jack-knifing[30].

[1]Center for Cell Decision Processes, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Departments of Biology and Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

of Adv–TNF synergy is specific to cell type: HeLa cervical carcinoma cells are much more sensitive to apoptosis than HT-29 colon adenocarcinoma cells when pre-infected with the same virus-to-cell ratio and stimulated with the same concentration of TNF[11] (Supplementary Fig. 1a, b).

There are two possible mechanisms by which cell-specific outcomes to a stimulus might be mediated. First, both the level of signal activation and the mechanism of effector processing could be cell specific (Fig. 1b). In this scenario, a given signalling event (for instance, the activation of a protein kinase) would contribute differently to a phenotype, depending on the particular cell type in which the signal was activated. A simpler alternative is that signal activation is sufficient to confer cell-specific outcomes (Fig. 1c). In this case, the main determinant of cell specificity would be the strength and combination of activation events induced by the stimulus. Different cell types could then use a common effector-processing mechanism that integrates these cell-specific signalling events to elicit cell-specific responses. Typically, the molecular details that link transducers to effectors and responses are incompletely understood. We therefore tested these two hypotheses by measuring the intracellular network and then mathematically calculating the bulk contribution of transducer proteins to effector processing (Fig. 1c; $f(x)$).

We selected an Adv–TNF treatment that causes a several-fold greater apoptotic response in HeLa cells than in HT-29 cells (Supplementary Fig. 1a–d). To monitor network activation induced by treatment with Adv and TNF, we measured the activities of five kinases that are important in Adv–TNF signalling and apoptosis: extracellular signal-regulated kinase (ERK), the serine-threonine protein kinase B (PKB)/Akt, c-Jun N-terminal kinase 1 (JNK1), inhibitor of nuclear factor (NF)-κB kinase (IKK), and MAPK-activated protein kinase 2 (MK2) (Supplementary Table 1). We observed significant differences in the dynamic activation of Akt, JNK1, IKK and MK2 between uninfected and Adv-infected HeLa cells that were stimulated with TNF ($P < 0.01$; Supplementary Fig. 2a), confirming that Adv had perturbed kinase signalling en route to apoptosis. Importantly, there were also significant differences in the dynamics of ERK, Akt, JNK1 and IKK activation between Adv-infected HT-29 cells and Adv-infected HeLa cells ($P < 0.001$; Supplementary Fig. 2b). Therefore, Adv infection perturbs the TNF-activated kinase network in epithelial cells[11] and the nature of the perturbation is cell specific.

Interferon-γ (IFN-γ) also sensitizes TNF-induced apoptosis and converges upon many of the same pathways as TNF and Adv[11,18]. Therefore, we compared the cell-specific activity measurements with published time courses of signalling induced by IFN-γ and TNF in HT-29 cells[18,19] (Fig. 1d). Hierarchical clustering of the concatenated kinase-activity time courses indicated that the HeLa cell signalling profiles, although different from HT-29 profiles for the same stimulus, were roughly similar to network states that could be achieved by HT-29 cells.

Because the signalling dynamics of HeLa and HT-29 cells were homologous overall (Fig. 1d) but their apoptotic responses clearly differed (Fig. 1d, Supplementary Fig. 1a–d), we could directly test the two effector-processing hypotheses (Fig. 1b, c). If the common-processing hypothesis were true, then a processing function (Fig. 1c; $f(x)$) deduced from data from HT-29 cells should accurately predict how activation events in HeLa cells are translated into HeLa-cell-specific apoptosis. Failure of such a model would argue that additional transducer pathways needed to be measured or that cell-specific processing was required (Fig. 1b).

We calculated the effector-processing function by using partial least squares regression (PLSR), a data-driven modelling approach that can connect cell outcomes to network activation[8,20]. We constructed a PLSR model from the dynamic kinase-activation profiles of the seven HT-29 treatment conditions shown in Fig. 1d (see Methods) and the corresponding apoptotic responses 24 and 48 h after TNF stimulation (Supplementary Fig. 1a, c). When the resulting model was then used *a priori* to predict apoptosis in HeLa cells, we found that the model predictions captured experimental values to within 92% (Fig. 1e). The remarkable accuracy of the HT-29 model for predicting HeLa-specific responses indicated that these two cell types might share a common effector-processing mechanism that converts network activation to apoptosis (Fig. 1c).

To exclude the possibility that common effector processing was unique to transformed cells, we used MCF-10A cells, an immortalized but non-transformed mammary epithelial cell line. Analysis of cell-cycle profiles for HT-29, HeLa and MCF-10A cells revealed that there were similar percentages of cells in G_1, S and G_2/M phase but indicated that the transformed lines carry $2N+$ DNA content owing to chromosomal aberrations (Supplementary Fig. 3a, Supplementary Table 2). Importantly, MCF-10A cells were highly resistant to apoptosis induced by treatment with Adv and TNF compared with HT-29 and HeLa cells (Supplementary Figs 1c, d, 3b), despite being responsive to TNF and readily infected by Adv (Supplementary Fig. 3c, d). To test whether the common effector-processing model could be extended to this third pattern of cell-specific apoptosis, we first stimulated MCF-10A cells with TNF and Adv plus TNF and measured the full time course of kinase activities. As expected, the signalling patterns stimulated by the two treatment combinations in MCF-10A cells were substantially different from those in the transformed lines (Fig. 1d). When the MCF-10A signalling data were input into the common effector-processing model, we found that the resistance of MCF-10A cells to apoptosis induced by treatment with Adv and TNF was correctly predicted within the confidence of the model and the data (Fig. 1e, Supplementary Fig. 4a). Thus, predictions of apoptosis by the common effector-processing model are not limited to transformed cells and might apply to epithelial cells in general.

The kinase activities that we measured are involved in many cellular phenotypes besides apoptosis. If common effector processing is a valid cell-biological principle, then similar models should be trainable for other cell-specific responses. We tested this idea by examining secretion of RANTES (also known as CCL5), a chemokine that is potently induced by both TNF[21] and infection with Adv[22]. The upstream promoter region of RANTES contains putative binding sites for several transcription factors that are direct substrates of the measured kinases[21]. Furthermore, we found that Adv significantly increased TNF-induced secretion of RANTES in HeLa and MCF-10A cells, but not in HT-29 cells (Supplementary Fig. 5). This cell-specific pattern of RANTES secretion was unlike that of apoptosis (Supplementary Figs 1a–d, 3b), indicating that the upstream signal processing was also distinct. Secretion of RANTES induced by treatment with Adv plus TNF therefore provided a cell-specific response that was plausibly controlled by the measured kinases but uncoupled from apoptosis.

We constructed and tested a new chemokine-secretion model by quantifying RANTES in the medium of HT-29, HeLa and MCF-10A cells, 12 and 24 h after treatment with the virus-cytokine combinations used earlier (Fig. 1d). We found that the HT-29 chemokine-secretion model accurately predicted cell-specific RANTES release for the three epithelial lines across five decades (Fig. 1f). HeLa and MCF-10A predictions were accurate to within 88% and 75%, respectively, with considerable overlap in the data-model confidence limits (Supplementary Fig. 4b). Thus, common effector processing applies to multiple cell-specific biological responses.

We next sought to determine whether any cell-specific insight could be gleaned directly from the common effector-processing function of the apoptosis model. The PLSR method derives a set of ‘principal components’, which contain linear combinations of the original kinase-activity measurements that are optimized to predict outcomes, such as apoptosis[20]. We assigned biological functions to the leading pair of principal components by looking at the projections of HT-29 cells treated with IFN-γ plus TNF. Compared to IFN-γ alone, the IFN-γ plus TNF conditions projected exclusively along

principal component 2 in a dose-dependent manner (Fig. 2a), indicating that principal component 2 was the axis of apoptotic synergy[8]. Compared to TNF alone, the IFN-γ plus TNF treatment also caused a decreased projection along principal component 1, which implied that this orthogonal axis reflected IFN-inducible responses that were unrelated to apoptosis. Consistent with this idea, we found that treatment with Adv (which does not cause apoptosis by itself but activates the type-I IFN response; see below) caused similar projections to treatment with IFN-γ. Consequently, we proposed that responsiveness to Adv–TNF treatment could be broken down into two parts that align with the principal components: an IFN response and an apoptotic response.

We quantified the relative proportions of the IFN and apoptotic responses by comparing the projections arising from treatment with TNF and Adv plus TNF for each cell type along the two principal component axes (Fig. 2b; Supplementary Fig. 6a). HeLa cells projected most strongly along the apoptosis axis, whereas MCF-10A cells had the greatest relative trajectory along the IFN axis, and HT-29 cells had roughly equal responses along both axes. The cell-specific projection along the apoptosis axis correlated with sensitivity to apoptosis induced by treatment with Adv plus TNF (Supplementary Figs 1c, d, 3b). To test the prediction of cell-specific IFN responses, we measured upregulation of the type-I *Ifna* and *Ifnb* genes, along with five IFN-induced genes that have been mechanistically implicated in the antiviral response[23] (Fig. 2c and Supplementary Fig. 6b). All three cell lines showed some induction of IFN expression in response to treatment with Adv, TNF or both. However, MCF-10A cells showed the most robust IFN response under conditions of Adv–TNF synergy, whereas HeLa cells showed the weakest IFN induction overall and completely lacked Adv–TNF synergy. Therefore, in addition to capturing cell-specific apoptotic responses, the common-processing model revealed an underlying cell-specific responsiveness along the IFN pathway.

An urgent application of common effector processing is to predict cell-specific responses to rational drug therapies[24]. We therefore tested whether the common-processing model could distinguish apoptotic outcomes between tumour cell types (HT-29 and HeLa) when kinase activities were perturbed by inhibitors. The phosphatidylinositol-3-OH kinase (PI(3)K)–Akt/PKB pathway is a clinically relevant cancer target[25] whose inhibition in Adv-infected HT-29 cells markedly increases TNF-induced cell death[11]. However, the role of PI(3)K–Akt signalling in TNF-induced apoptosis of Adv-infected HeLa cells has been unknown.

To predict the effects of inhibition of PI(3)K on HeLa cell viability, we updated the model with measured values[11] of Akt inhibition and apoptosis in HT-29 cells treated with the PI(3)K inhibitor LY294002. We assumed that the potency of LY294002 was constant throughout the 24 h after TNF addition and that no other kinases were affected (see Methods). Next, we experimentally measured the extent to which LY294002 reduced Akt activity in HeLa cells (Fig. 3a). Combining the LY294002-inhibited Akt activity with the corresponding HeLa-network measurements (Fig. 1d), we then used the updated model to predict the response to inhibition of the PI(3)K–Akt pathway when HeLa cells were treated with Adv plus TNF. Surprisingly, the updated model predicted no significant increase

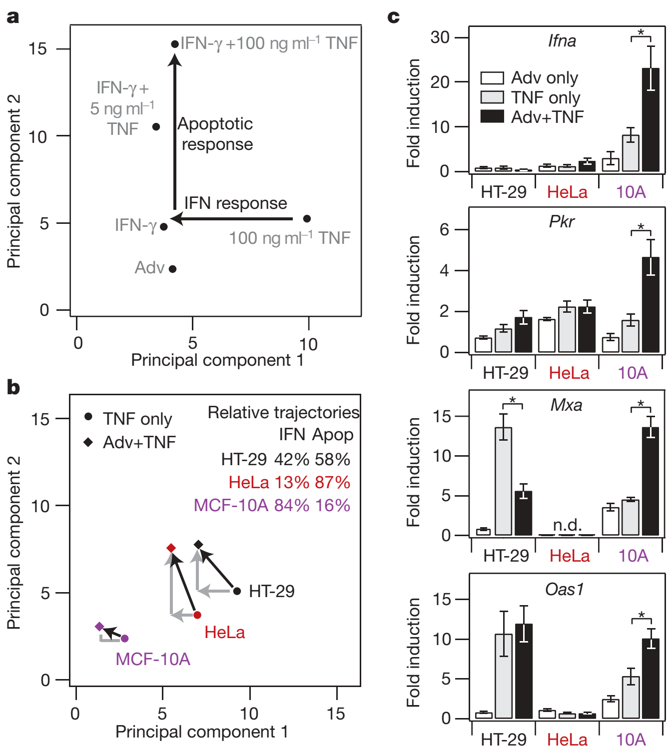


Figure 2 | Principal components of the common-processing model reveal cell-specific IFN responses to treatment with Adv plus TNF. **a**, **b**, Reduced principal component dimensions of the common-processing model for HT-29 cells treated with IFN-γ, Adv and TNF (**a**); and HT-29 (black), HeLa (red) and MCF-10A (purple) cells treated with TNF alone (circles) or Adv+TNF (diamonds) (**b**). Protein-kinase signals for each component are included in Supplementary Tables 3 and 4. Apop, apoptosis. **c**, Measurements of IFN-gene induction across cell types. HT-29, HeLa and MCF-10A cells were treated as indicated and collected 16 h later. Experimental data are presented as the mean ± s.e. ($n = 3$). *Pkr*, gene for double-stranded RNA-dependent protein kinase; *Mxa*, gene for Mx protein GTPase; *Oas1*, gene for 2′,5′-oligoadenylate synthetase 1.

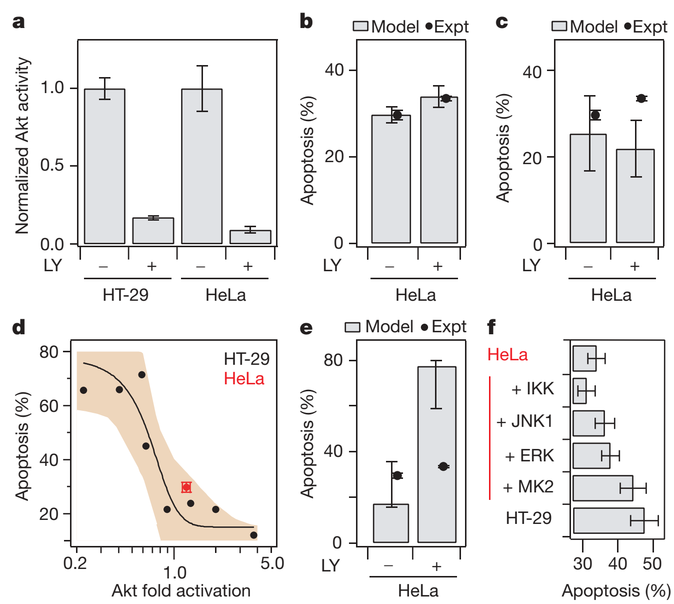


Figure 3 | Common effector processing uniquely predicts resistance of Adv-infected HeLa cells to PI(3)K inhibition. **a**, Akt activity measured in Adv-infected HT-29 cells and HeLa cells 12 h after TNF stimulation with or without 20 μM LY294002 (LY). Experimental data are presented as the mean ± s.e. ($n = 6$ for HT-29, $n = 6$ for HeLa). **b**, **c**, Apoptosis predicted by the common-processing model (**b**) or HeLa cell-specific processing model (**c**) compared with that measured experimentally (expt) in HeLa cells treated with Adv plus TNF with or without 20 μM LY. **d**, One-kinase relationship linking Akt activity to apoptosis of HT-29 cells (black) treated with combinations of IFN-γ, Adv, and TNF[11] and HeLa cells (red) treated with 100 MOI Adv and 100 ng ml^{-1} TNF. 95% confidence interval (tan) is shown. **e**, Apoptosis predicted by the Akt-apoptosis model and measured experimentally (markers) in HeLa cells treated with Adv plus TNF with or without 20 μM LY. **f**, Apoptosis predicted by the common-processing model for HT-29–HeLa hybrid-cell profiles containing mixtures of HT-29 signals (black) and HeLa signals (red). Experimental data are presented as the mean ± s.e. ($n = 3$), and model uncertainties are estimated by jack-knifing[30].

in apoptosis in HeLa cells treated with Adv plus TNF after inhibition of PI(3)K activity, which we confirmed by direct measurement in PI(3)K-inhibited HeLa cells (Fig. 3b and Supplementary Fig. 7). Thus, both the model and experiment indicate that the role of PI(3)K in apoptosis induced by treatment with Adv plus TNF is cell specific: apoptosis of HT-29 cells is PI(3)K-sensitive[11], whereas apoptosis of HeLa cells is PI(3)K-resistant.

The accurate LY294002 predictions in HeLa cells provided us with the opportunity to compare the common model directly against competing mechanisms of effector processing. To mimic cell-specific processing (Fig. 1b), we trained a separate function based only on HeLa measurements and tested whether the existing HeLa data were superior for correctly relating signals to apoptosis. With the relatively few observations, the HeLa-specific model was highly uncertain and far less accurate than the common-processing model based on HT-29 data (Fig. 3c). This emphasized that HT-29-specific information, and thus common effector processing, was essential for prediction using the present data set.

A reductionist alternative to common effector processing is that there exists one crucial transducer, which is a master regulator of Adv–TNF synergy across cell types. TNF-induced apoptosis is highly dependent on the relative extent of Akt activation[11,26], and Akt activation 12 h after treatment predicted apoptosis in Adv-infected HeLa cells (Fig. 3d). This raised the possibility that Akt signalling alone was sufficient to predict cell-specific apoptosis. However, when the Akt model was used with LY294002-inhibited measurements, it incorrectly predicted a fourfold increase in apoptosis in HeLa cells treated with Adv plus TNF (Fig. 3e), indicating the need for signalling information from multiple pathways. This requirement was investigated further by individually substituting HT-29-cell kinase-activity time courses into the LY294002-inhibited HeLa cell data set to construct hybrid-cell profiles (see Methods). In the common-processing model, no one kinase was dominant (Fig. 3f), in agreement with the distributed information provided by these pathways (Supplementary Fig. 8). Among the competing mechanisms for achieving cell specificity, we conclude that common effector processing is most consistent with the experimental data.

The IKK–NF-κB pathway is another important drug target for treating both cancer and inflammatory diseases[27]. However, the value of IKK or NF-κB inhibitors has been debated because the pathway operates very differently across cell types[5]. We used the common-processing model to predict how inhibition of the IKK–NF-κB pathway by SC-514 would affect apoptosis in HT-29 and HeLa cells. We measured the extent to which SC-514 inhibited IκBα degradation in HT-29 and HeLa cells (Supplementary Fig. 9) and then simulated IKK inhibition as described for Akt (see Methods).

Given the new SC-514 signalling inputs, the common-processing model predicted that pretreatment of HT-29 cells with SC-514 would negligibly affect apoptosis (Fig. 4a upper; conditions 1–4), which almost exactly matched the measured values (Fig. 4a upper; conditions 5–8). By contrast, apoptosis of HeLa cells was predicted to increase substantially in uninfected cells treated with SC-514 (Fig. 4a lower; conditions 1–2) and to decrease slightly in Adv-infected cells (Fig. 4a lower; conditions 3–4). Validation experiments showed that the model roughly captured the SC-514-mediated increase in TNF-induced apoptosis of uninfected HeLa cells (Fig. 4a lower; conditions 1–2 and 5–6) but incorrectly predicted apoptosis after SC-514 inhibition in Adv-infected HeLa cells (Fig. 4a lower; conditions 3–4 and 7–8). We found the same pattern of cell-specific apoptosis when using two other small-molecule inhibitors that target different points on the IKK–NF-κB pathway (Supplementary Fig. 10). Thus, the common-processing model had qualitatively predicted the existence of cell-specific responses to inhibition of the IKK–NF-κB pathway, but did not quantitatively predict the extent of specificity in HeLa cells treated with Adv plus TNF.

The discrepancy between the model and the experimental data was not necessarily due to failure of common processing but instead could arise from limitations in the HT-29 data that were used to train the PLSR model[20]. We compared the distribution of IKK measurements in HeLa cells with the HT-29 training data and found that the data for treatment with Adv plus TNF contained several IKK-activity observations that were higher than any previously measured in HT-29 cells (Supplementary Fig. 11). Furthermore, all of these very high IKK activities occurred within 4 h of TNF treatment (Fig. 4b), when IKK activity is considered to be anti-apoptotic[8,27]. These observations indicated that the HT-29 training set was not diverse enough to quantify the common-effector relationship between early IKK activity and apoptosis. Predictions of HeLa apoptosis when IKK was inhibited would therefore be uncertain.

An important corollary to this explanation is that early IKK activity should have a disproportionately anti-apoptotic role in HeLa cells. We tested this by performing timed-inhibition experiments with SC-514. Direct inhibition of early IKK signalling was not feasible, because removal of SC-514 caused hyperactivation of IKK (Supplementary Fig. 12). Therefore, we selectively inhibited IKK from 4 h to 24 h and compared the increase in apoptosis to when IKK was inhibited from 0 h to 24 h (Fig. 4c). In HT-29 cells, IKK inhibition from 4 h to 24 h had the same effect as inhibiting IKK from 0 h to 24 h. By contrast, nearly half of the SC-514-mediated increase in HeLa-cell apoptosis could be attributed to early-phase inhibition of IKK (Fig. 4c and Supplementary Fig. 13). Early IKK activation therefore appears to be specifically crucial for anti-apoptotic signalling in Adv-infected HeLa cells. Together, these data support common effector processing as part of the mechanism for achieving cell-specific sensitivity to pharmacological inhibitors and apoptotic stimuli.

Our study reveals an underlying similarity among disparate cell types that becomes evident when signalling information is processed at the network level. Most of the kinases measured here lie at the 'waist' of hourglass-shaped pathways[28], indicating a unique point of convergence in the network architecture (Fig. 1a) that can be exploited to make predictions broadly and accurately. The predictions of cell-specific apoptosis, chemokine secretion, IFN induction and drug sensitivity were focused on epithelial cells for the sake of feasibility. There probably does not exist a common-processing

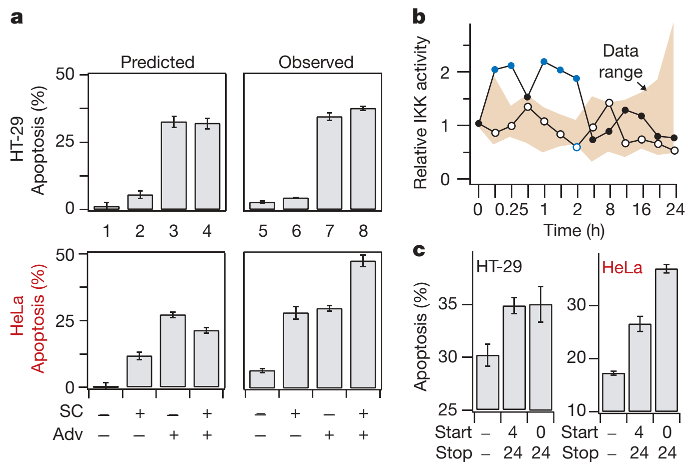


Figure 4 | Deduction of cell-specific sensitivity to early-phase IKK inhibition. **a**, Apoptosis predicted by the common-processing model (left) and measured experimentally (right) in HT-29 (top) and HeLa (bottom) cells treated with 100 ng ml^{-1} TNF with or without 20 μM SC-514 (SC) and 100 MOI (HeLa) or 1,000 MOI (HT-29) Adv pre-infection. **b**, Range plot of HT-29 IKK activity used for model training (tan) and IKK activity time course of HeLa cells treated with TNF (open) or Adv plus TNF (filled). Points outside the training range are highlighted in blue. **c**, TNF-induced apoptosis in Adv-infected HT-29 and HeLa cells following IKK inhibition with 20 μM SC-514 for the indicated times. Data are presented as the mean ± s.e. ($n = 3$), and model uncertainties are estimated by jack-knifing[30].

function that captures responses across all cell types. Consistent with this expectation, our preliminary attempts to extend a reduced version of the common effector-processing model to apoptosis of Jurkat cells (a leukaemic T-cell line) led to highly uncertain predictions (Supplementary Fig. 14). Nonetheless, for a diverse range of epithelia, our data support the existence of common effector mechanisms that can predict cell fate. Such mechanisms are revealed only after applying a data-driven modelling approach that converts important kinase-signalling events into a phenotype[8,20].

If the common-processing hypothesis had failed, it would have meant that each cell type would require its own in-depth experiments and model training. Our results raise the possibility that models can be refined with new data to the point that they enable accurate predictions of phenotype that are broadly applicable. Already, the distinction between two carcinoma lines on the basis of their apoptotic response to PI(3)K and IKK inhibition points towards immediate applications for combining pharmacological inhibitors[6] and cancer gene therapy[7] to treat certain tumours.

METHODS SUMMARY

HT-29 cells, HeLa cells (American Type Culture Collection) and MCF-10A cells were grown as described[11,29]. Infection with Adv, stimulation with TNF, pretreatment with IFN-γ and lysis were performed essentially as described[11,18]. Small molecule inhibitors (Calbiochem) were spiked into media 1–2 h before TNF stimulation. Kinase activities were measured by a microtitre-based assay[26] and were confirmed to operate in the linear range for each cell type (Supplementary Fig. 15). Values are normalized to the 0-min time point and then multiplied by cell-specific and kinase-specific correction factors (Supplementary Table 5). Descriptors of signalling dynamics were derived as described elsewhere[8]. Network-activation time courses were concatenated, log transformed, and normalized to the maximum observed value for each kinase. Clustering was performed by the unweighted pair group method with arithmetic mean and a Euclidean distance metric. Apoptosis was measured by flow cytometry for cleaved caspase-cytokeratin as described[11]. For RANTES secretion, we collected supernatants and analysed them according to the manufacturer's recommendations (R&D Systems). Gene induction was measured by real-time quantitative PCR on a LightCycler II instrument (Roche) with standard curves based on purified amplicons. The PLSR model was constructed using SIMCA-P (Umetrics) as described[8]. Model calibrations were made by leave-one-out cross-validation[19] for the HT-29 treatments, and model uncertainties were calculated by jack-knifing[30]. For inhibitor predictions, the target kinase activity was held constant at the experimentally measured inhibited value, time-course activities of the other kinases were kept the same as before, and signalling descriptors were recalculated. We used Student's *t*-test to compare two individual means, and two-factor ANOVA to compare two activity time courses. We calculated nonlinear confidence intervals by support plane analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 April; accepted 7 June 2007.
Published online 18 July 2007.

1. Gerhart, J. 1998 Warkany lecture: signaling pathways in development. *Teratology* 60, 226–239 (1999).
2. Jordan, J. D., Landau, E. M. & Iyengar, R. Signaling networks: the origins of cellular multitasking. *Cell* 103, 193–200 (2000).
3. Downward, J. The ins and outs of signalling. *Nature* 411, 759–762 (2001).
4. Irish, J. M. *et al.* Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 118, 217–228 (2004).
5. Wajant, H., Pfizenmaier, K. & Scheurich, P. Tumor necrosis factor signaling. *Cell Death Differ.* 10, 45–65 (2003).
6. Kenakin, T. Predicting therapeutic value in the lead optimization phase of drug discovery. *Nature Rev. Drug Discov.* 2, 429–438 (2003).
7. McCormick, F. Cancer gene therapy: fringe or cutting edge? *Nature Rev. Cancer* 1, 130–141 (2001).
8. Janes, K. A. *et al.* A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 310, 1646–1653 (2005).
9. Schmitz, E. K., Kraus, D. M. & Bulla, G. A. Tissue-specificity of apoptosis in hepatoma-derived cell lines. *Apoptosis* 9, 369–375 (2004).
10. Ohmori, Y. & Hamilton, T. A. Cell type and stimulus specific regulation of chemokine gene expression. *Biochem. Biophys. Res. Commun.* 198, 590–596 (1994).
11. Miller-Jensen, K., Janes, K. A., Wong, Y. L., Griffith, L. G. & Lauffenburger, D. A. Adenoviral vector saturates Akt pro-survival signaling and blocks insulin-mediated rescue of tumor necrosis-factor-induced apoptosis. *J. Cell Sci.* 119, 3788–3798 (2006).
12. Pawson, T. Specificity in signal transduction: from phosphotyrosine-sh2 domain interactions to complex cellular systems. *Cell* 116, 191–203 (2004).
13. Bouwmeester, T. *et al.* A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. *Nature Cell Biol.* 6, 97–105 (2004).
14. Cho, K. H., Shin, S. Y., Lee, H. W. & Wolkenhauer, O. Investigations into the analysis and modeling of the TNFα-mediated NF-κB-signaling pathway. *Genome Res.* 13, 2413–2422 (2003).
15. Janes, K. A. & Lauffenburger, D. A. A biological approach to computational models of proteomic networks. *Curr. Opin. Chem. Biol.* 10, 73–80 (2006).
16. Philpott, N. J., Nociari, M., Elkon, K. B. & Falck-Pedersen, E. Adenovirus-induced maturation of dendritic cells through a PI3 kinase-mediated TNF-alpha induction pathway. *Proc. Natl Acad. Sci. USA* 101, 6200–6205 (2004).
17. Liu, T. C. *et al.* Functional interactions of antiapoptotic proteins and tumor necrosis factor in the context of a replication-competent adenovirus. *Gene Ther.* 12, 1333–1346 (2005).
18. Janes, K. A. *et al.* The response of human epithelial cells to TNF involves an inducible autocrine cascade. *Cell* 124, 1225–1239 (2006).
19. Gaudet, S. *et al.* A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol. Cell. Proteomics* 4, 1569–1590 (2005).
20. Janes, K. A. & Yaffe, M. B. Data-driven modelling of signal-transduction networks. *Nature Rev. Mol. Cell Biol.* 7, 820–828 (2006).
21. Nelson, P. J., Kim, H., Manning, W., Goralski, T. & Krensky, A. Genomic organization and transcriptional regulation of the RANTES chemokine gene. *J. Immunol.* 151, 2601–2612 (1993).
22. Bowen, G. P. *et al.* Adenovirus vector-induced inflammation: capsid-dependent induction of the C–C chemokine RANTES requires NF-κB. *Hum. Gene Ther.* 13, 367–379 (2002).
23. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* 14, 778–809 (2001).
24. Dancey, J. & Sausville, E. A. Issues and progress with protein kinase inhibitors for cancer treatment. *Nature Rev. Drug Discov.* 2, 296–313 (2003).
25. Luo, J., Manning, B. D. & Cantley, L. C. Targeting the PI3K-Akt pathway in human cancer: rationale and promise. *Cancer Cell* 4, 257–262 (2003).
26. Janes, K. A. *et al.* A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis. *Mol. Cell. Proteomics* 2, 463–473 (2003).
27. Karin, M. & Greten, F. R. NF-κB: linking inflammation and immunity to cancer development and progression. *Nature Rev. Immunol.* 5, 749–759 (2005).
28. Csete, M. & Doyle, J. Bow ties, metabolism and disease. *Trends Biotechnol.* 22, 446–450 (2004).
29. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 30, 256–268 (2003).
30. Janes, K. A. *et al.* Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J. Comput. Biol.* 11, 544–561 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Hoffman, K. Haigis and members of the Lauffenburger and Brugge laboratories for comments on the manuscript. This work was supported by grants from the NIGMS Cell Decision Processes Center, the USCB-CalTech-MIT Institute for Collaborative Biotechnologies, and the MIT Biotechnology Process Engineering Center to D.A.L. K.A.J. acknowledges support from the American Cancer Society (New England Division – SpinOdyssey).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.L. (lauffen@mit.edu).

METHODS

Adenoviral vectors. The recombinant adenovirus type 5 vectors were deleted for E1 and E3 regions and expressed *Escherichia coli* β-galactosidase (β-gal) under control of the cytomegalovirus (CMV) enhancer/promoter (provided by the University of Michigan Vector Core). Virus was reported to have an infectious plaque-forming unit (p.f.u.) concentration of 1 p.f.u. for every 23.5 viral particles (v.p.) as determined by plaque assay.

Cell culture and stimulation. HT-29 and HeLa cells were grown according to ATCC recommendations and plated at 50,000 and 20,000 cells per cm^2, respectively. The cell densities for HT-29 and HeLa cells were optimized to have approximately the same number of cells at the end of the standardized 4-day experimental protocol. MCF-10A cells were grown as described[29] and plated at a density of 50,000 cells per cm^2. HT-29 cells were infected with 0, 1.4×10^9 or 1.4×10^{10} v.p. ml^{-1} Adv (0, 100 or 1,000 MOI), HeLa cells with 0 or 7.4×10^8 v.p. ml^{-1} Adv (0 or 100 MOI), and MCF-10A cells with 0 or 1.4×10^9 v.p. ml^{-1} Adv (0 or 100 MOI). Infections were performed essentially as described[11]. The infection MOI for HT-29 and HeLa cells was chosen to maximize the difference in cell response according to published Adv dose responses in the presence of 100 ng ml^{-1} TNF[11]. Cells pretreated with IFN-γ were prepared and lysed as described[18]. Twenty-four hours after the start of infection, cells were spiked with 100 ng ml^{-1} TNF (Peprotech) for 5, 15, 30, 60 and 90 min and 2, 4, 8, 12, 16, 20 and 24 h and lysed as described[11,26]. For Akt perturbation, LY294002 or wortmannin (Calbiochem) was spiked into the medium 1 h before TNF stimulation. For IKK–NF-κB pathway perturbation, SC-514, SN-50, or BAY 11-1082 (Calbiochem) was spiked into the medium 2 h before TNF stimulation. For small-molecule inhibitor controls, the carrier alone (0.02–0.1% DMSO) or SN50M (control peptide for SN50) was added.

Network-level signalling measurements. Kinase activities for five network signals—ERK, Akt, JNK1, IKK and MK2—were quantitatively measured by a microtitre-based assay[26]. All kinase assays were confirmed to operate in the linear range for each cell type (Supplementary Fig. 15). Values are presented as the mean of duplicate biological samples, normalized to the 0-min time point from quadruplicate biological samples. To adjust for cell-specific differences in basal activities, kinase-specific correction factors were calculated for each cell type by scaling the median 0-min (untreated) activity values (Supplementary Table 5) to the 0-min activity of the kinase in HT-29 cells (assigned a nominal value of 1). The fold activations of HeLa and MCF-10A cells were then multiplied by the cell- and kinase-specific correction factors to fuse the data sets for the PLSR-model predictions. The final processed data were normalized to the maximum value of each signal across all conditions to aid comparison of signals with different dynamic ranges. Descriptors of signalling dynamics were derived for each time course as described[8].

Apoptosis measurements. Cells that scored double-positive for cleaved cytokeratin (a caspase-3/6/7 substrate) and cleaved-caspase-3 were measured by flow cytometry as described[11].

Enzyme-linked immunosorbant assay. HT-29, HeLa and MCF-10A cells were treated in parallel according to the conditions described in Fig. 1d. Supernatants were collected after 12 and 24 h and analysed for RANTES secretion according to the manufacturer's recommendations (R&D Systems). Receptors for RANTES are not detectable in HT-29[31], HeLa[32] and MCF-10A cells (K.A.J. and J.S.B., unpublished observations), which avoids the confounding effects of autocrine binding in these cells. Measurements were normalized to live-cell counts collected in parallel with the supernatants after 12 and 24 h. To account for different basal levels of secretion across cell types, the HT-29 secretion data were mean-centred and variance-scaled before training the PLSR model. The predicted TNF-induced RANTES secretion in HeLa and MCF-10A cells were transformed back to absolute physical units based on the mean and variance of observed HeLa and MCF-10A RANTES secretion data across four conditions (no treatment, TNF only, Adv plus TNF, and IFN-γ plus TNF).

Real-time quantitative PCR (RT-qPCR). HT-29, HeLa and MCF-10A cells were seeded in 12-well plates and treated with Adv and TNF as described. Sixteen hours after TNF stimulation, RNA was prepared by lysing in 0.5 ml RNA STAT-60 (Tel-Test). After phenol-chloroform purification, residual genomic DNA was eliminated with DNA-free (Ambion). First-strand synthesis was performed with 5 μg DNA-free RNA, Superscript III (Invitrogen), and oligo(dT)24 according to the manufacturer's recommendations. RT-qPCR was done using 0.1 μl cDNA template and 10 pmol forward and reverse primers together with a homemade master mix used at a final concentration of 1 × PCR Buffer II (Applied Biosystems), 4 mM $MgCl_2$, 200 μM each of dATP, dCTP, dGTP and dTTP, 150 μg ml^{-1} BSA, 5% glycerol, 0.25 × SYBR green (Invitrogen), and 0.025 U ml^{-1} AmpliTaq (Applied Biosystems) in a final reaction volume of 15 μl. Real-time reactions were tracked on a LightCycler II instrument (Roche) with the following amplification scheme: a denaturation step of 95 °C for 90 s; an amplification step of 95 °C for 10 s, 60 °C for 10 s, 72 °C for 12 s repeated 40 times; a fusion step of 65 °C to 95 °C touched up at 0.1 °C s^{-1}. All primers gave single peaks upon melting-curve analysis (sequences are listed in Supplementary Table 6). Relative levels were determined on the basis of a standard curve of gene-specific amplicons diluted through five orders of magnitude. Samples were normalized to the geometric mean of three housekeeping genes: glyceraldehyde-3-dehydrogenase (*gapdh*), peroxiredoxin 6 (*prdx6*) and histidine triad nucleotide binding protein 1 (*hint1*) (Supplementary Table 6)[33]. Fold induction was calculated for each cell type by normalizing to its own mock-stimulation levels.

Hierarchical clustering. Network-activation time courses were concatenated, log transformed, and normalized to the maximum observed value for each kinase. Clustering was performed by the unweighted pair group method with arithmetic mean and a Euclidean distance metric.

PLSR model construction and refinement. The PLSR model was constructed using SIMCA-P (Umetrics) as described[8]. The predictive power of the HT-29 model was calculated by leave-one-out cross-validation[19] (73% predictive power for apoptosis model and 81% for RANTES secretion). Model uncertainties were calculated by jack-knifing[30]. Treatment projections were plotted after a 109° subspace rotation of the principal component vectors. To approximate the dynamic network behaviour for a given treatment in the presence of PI(3)K inhibition, we held the level of Akt constant at the LY294002-inhibited activity fraction measured after 12 h (Fig. 3a) for HT-29 (0.17) and HeLa (0.09) compared to a 0-min baseline. Similarly, IKK activity levels were held constant at the measured SC-514-inhibited activity fraction (Supplementary Fig. 8; 0.65 for HT-29; 0.48 for HeLa). With both inhibitor predictions, time-course values of the other kinases were kept identical to the network behaviour measured in the absence of inhibitor, and metrics were extracted as described above. For HT-29–HeLa cell hybrid profiles, the measured HeLa kinase activities from Adv–TNF treatment were replaced one at a time with the corresponding kinase activities measured in HT-29 cells, and metrics were extracted as described above.

Statistical analysis. We used Student's *t*-test to compare two means, and two-factor ANOVA to compare two activity time courses. All tests were performed at a significance level of $\alpha = 0.05$. Nonlinear confidence intervals were calculated by support plane analysis[34].

31. Jordan, N. J. *et al.* Expression of functional CXCR4 chemokine receptors on human colonic epithelial cells. *J. Clin. Invest.* **104**, 1061–1069 (1999).

32. Choe, H. *et al.* The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135–1148 (1996).

33. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034.1–0034.11 (2002).

34. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* 2nd edn 118–124 (Kluwer Academic/Plenum, New York, 1999).

Vol 448 | 2 August 2007 | doi:10.1038/nature05936

LETTERS

Crystal structure of a human membrane protein involved in cysteinyl leukotriene biosynthesis

Hideo Ago[1]*, Yoshihide Kanaoka[2]*, Daisuke Irikura[1,2]*, Bing K. Lam[2], Tatsuro Shimamura[1], K. Frank Austen[2] & Masashi Miyano[1]

The cysteinyl leukotrienes, namely leukotriene (LT)C_4 and its metabolites LTD_4 and LTE_4, the components of slow-reacting substance of anaphylaxis[1,2], are lipid mediators of smooth muscle constriction[3–5] and inflammation[6,7], particularly implicated in bronchial asthma[8,9]. LTC_4 synthase (LTC_4S), the pivotal enzyme for the biosynthesis of LTC_4 (ref. 10), is an 18-kDa integral nuclear membrane protein[11,12] that belongs to a superfamily of membrane-associated proteins in eicosanoid and glutathione metabolism that includes 5-lipoxygenase-activating protein, microsomal glutathione *S*-transferases (MGSTs), and microsomal prostaglandin E synthase 1 (ref. 13). LTC_4S conjugates glutathione to LTA_4, the endogenous substrate derived from arachidonic acid through the 5-lipoxygenase pathway[14]. In contrast with MGST2 and MGST3 (refs 15, 16), LTC_4S does not conjugate glutathione to xenobiotics[17]. Here we show the atomic structure of human LTC_4S in a complex with glutathione at 3.3 Å resolution by X-ray crystallography and provide insights into the high substrate specificity for glutathione and LTA_4 that distinguishes LTC_4S from other MGSTs. The LTC_4S monomer has four transmembrane α-helices and forms a threefold symmetric trimer as a unit with functional domains across each interface. Glutathione resides in a U-shaped conformation within an interface between adjacent monomers, and this binding is stabilized by a loop structure at the top of the interface. LTA_4 would fit into the interface so that Arg 104 of one monomer activates glutathione to provide the thiolate anion that attacks C6 of LTA_4 to form a thioether bond, and Arg 31 in the neighbouring monomer donates a proton to form a hydroxyl group at C5, resulting in 5(*S*)-hydroxy-6(*R*)-*S*-glutathionyl-7,9-*trans*-11,14-*cis*-eicosatetraenoic acid (LTC_4). These findings provide a structural basis for the development of LTC_4S inhibitors for a proinflammatory pathway mediated by three cysteinyl leukotriene ligands whose stability and potency are different and by multiple cysteinyl leukotriene receptors whose functions may be non-redundant.

Recombinant human LTC_4S was crystallized with glutathione (GSH) and the structure was determined by the multiwavelength anomalous diffraction method by using the diffraction images from native LTC_4S and a selenomethionine (SeMet)-substituted Leu121Met mutant (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). The X-ray crystallography shows that LTC_4S is a homotrimer as predicted by two-dimensional crystallization and electron crystallographic studies for LTC_4S (ref. 18) as well as MGST1 (ref. 19) and microsomal prostaglandin E synthase 1 (ref. 20). In a LTC_4S monomer there are five α-helices, of which four are transmembrane helices (helices I, II, III and IV); the other, helix V, extrudes into the solvent (Fig. 1a). Both the amino and carboxy termini of a monomer exist on the same side of the membrane as predicted[21]. Because there are kinks of about 30° both in helix II at the middle and in helix IV after a short segment, helix II is subdivided into IIa and IIb, and helix IV into IVa and IVb.

GSH is bound in a cavity formed at the interface between neighbouring monomers. For simplicity, monomers C and A are used as an example of the neighbouring monomers to describe the active site

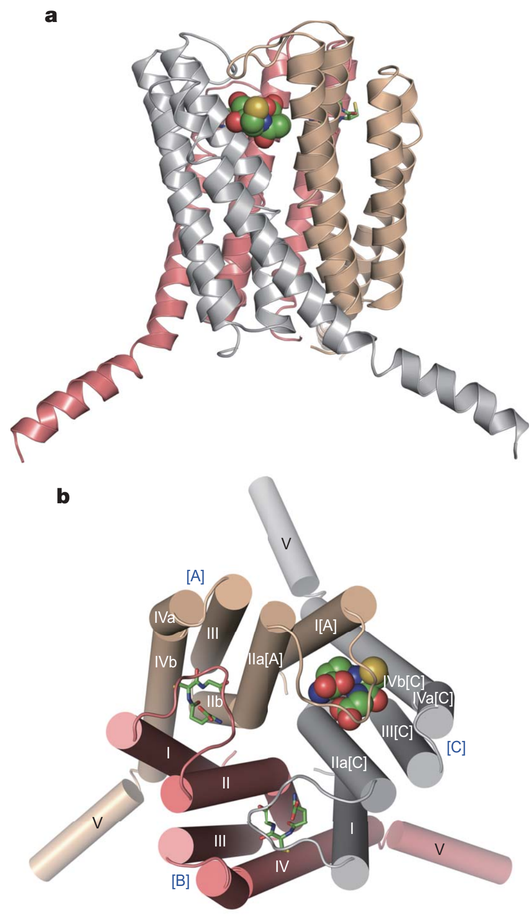


Figure 1 | Structure of LTC_4S. **a**, Side view of a LTC_4S homotrimer with bound GSHs. **b**, Top view of LTC_4S trimer (monomers A to C) with GSHs from lumen side. Five α-helices (I to V) with subsegments (a, b) in helix II and IV are shown. One GSH at the interface of [C] and [A] is shown in a space-filling model; the others are shown in a stick model.

[1]Structural Biophysics Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan. [2]Department of Medicine, Harvard Medical School and Division of Rheumatology, Immunology, and Allergy, Brigham Women's Hospital, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

and the key residues. As shown in Fig. 1b, the shallow concave formed by helices IIa, IIb, III and IVa and IVb in monomer C is located adjacent to helices I and IIa of the next monomer A, resulting in a GSH-binding site. The top of the GSH-binding site is covered by a loop between helix I and II (loop I–II) of monomer A. The three embedded GSH molecules in the cavities indicate that the LTC_4S trimer may be a functional unit for the conjugation of GSH to LTA_4 to form LTC_4.

The key residues for the GSH-binding site include Arg 51, Asn 55, Glu 58, Tyr 59, Tyr 93, Tyr 97 and Arg 104 of monomer C, and Arg 30 and Gln 53 of monomer A (Fig. 2). The bound GSH is in a U-shaped conformation, forming an intramolecular hydrogen bond between the backbone nitrogen atom of the glycyl moiety and an α-carboxyl oxygen of the γ-glutamyl moiety. This is quite distinct from the findings that a bound GSH observed in cytosolic GSTs[22] and probably in MGST1 (ref. 19) is in an extended conformation.

Neutralization of the charges on the GSH by the amino-acid residues from both neighbouring LTC_4S monomers would stabilize the U-shaped conformation beyond the coulombic repulsion between the carboxyl groups of the GSH (Fig. 2b). In the bound GSH, the α-carboxyl group of the γ-glutamyl and the C-terminal carboxyl group of the glycyl moieties form the salt bridges to Arg 30 of monomer A and Arg 51 of monomer C, respectively. The positively charged amine group of the γ-glutamyl moiety is countered by the carboxyl group of Glu 58 of monomer C. Furthermore, the charged guanidino group of Arg 104 of monomer C interacts with the thiol group of the cysteinyl moiety in the GSH.

The importance of Arg 51 in the binding of GSH is supported by the finding that LTC_4S mutants in which Arg 51 was replaced by isoleucine or threonine lost their LTC_4S activity[23]. The requirement of a positive charge at residue 51 was predicted by the finding that an Arg51His mutation retained the activity. Both of the arginine residues corresponding to Arg 30 and Arg 51 of LTC_4S are conserved between the members of the superfamily of membrane-associated proteins in eicosanoid and GSH metabolism that possess GST activity, whereas these residues are not conserved in 5-lipoxygenase-activating protein, which lacks GST activity[13] (Supplementary Fig. 2). The side chain of Arg 104 is presumed to abstract a proton[24], thus activating the bound GSH to form the thiolate anion, the reactive species in the GSH transfer catalysis. This is in a striking contrast to cytosolic GSTs, in which a hydroxyl group of Tyr or Ser activates GSH[22]. Arg 104 is preserved in MGST2 and conservatively substituted in MGST3, but is not conserved in 5-lipoxygenase-activating protein[13] (Supplementary Fig. 2). Because the resolution of the LTC_4S crystal structure is 3.3 Å, we cannot exclude the possibility that water molecules could be involved in activating GSH. Nonetheless, the three arginine residues identified as Arg 30, Arg 51 and Arg 104 seem to have an essential function in binding and catalysing the conjugation between the substrates GSH and LTA_4.

Site-directed mutagenesis of human LTC_4S had suggested that Tyr 93 was involved in GSH activation[23]. In addition, mutations of Asn 55, Tyr 59 and Tyr 97 did not affect the enzyme's activity but substantially increased its K_m for GSH. The crystal structure reveals that these residues contribute to the binding of GSH through the formation of hydrogen bonds (Fig. 2b, c).

The bound GSH molecule is separated from the bulk solvent by loop I–II; the loop is stabilized by the hydrogen-bond network between the carbonyl oxygens of Val 35, Pro 37 and Leu 39 of loop I–II of monomer A and the carboxyl group of the bound GSH through the side chain of Arg 30 of monomer A (Fig. 2c). In monomers without bound GSH, loop I–II would be more mobile so that GSH molecules in bulk solvent can bind to the GSH-binding site through the opened entrance.

The closure of the loop by the binding of GSH before LTA_4 would be appropriate for catalysing the conjugation of GSH to LTA_4. The reactive epoxide of LTA_4 can be hydrolysed by water at physiological pH with a half-time of a few seconds at 37 °C (ref. 25). As a result of the prior binding of GSH, the thiol group is separated from the bulk solvent, allowing conjugation with LTA_4 to proceed in the hydrophobic environment. If LTA_4 encountered LTC_4S without prior GSH binding, the reactive epoxide would be exposed to the aqueous environment and would undergo hydrolysis to four non-enzymatic products[26].

In the crystal LTC_4S structure, a molecule of n-dodecyl-β-D-maltoside (DDM), the detergent used to solubilize recombinant

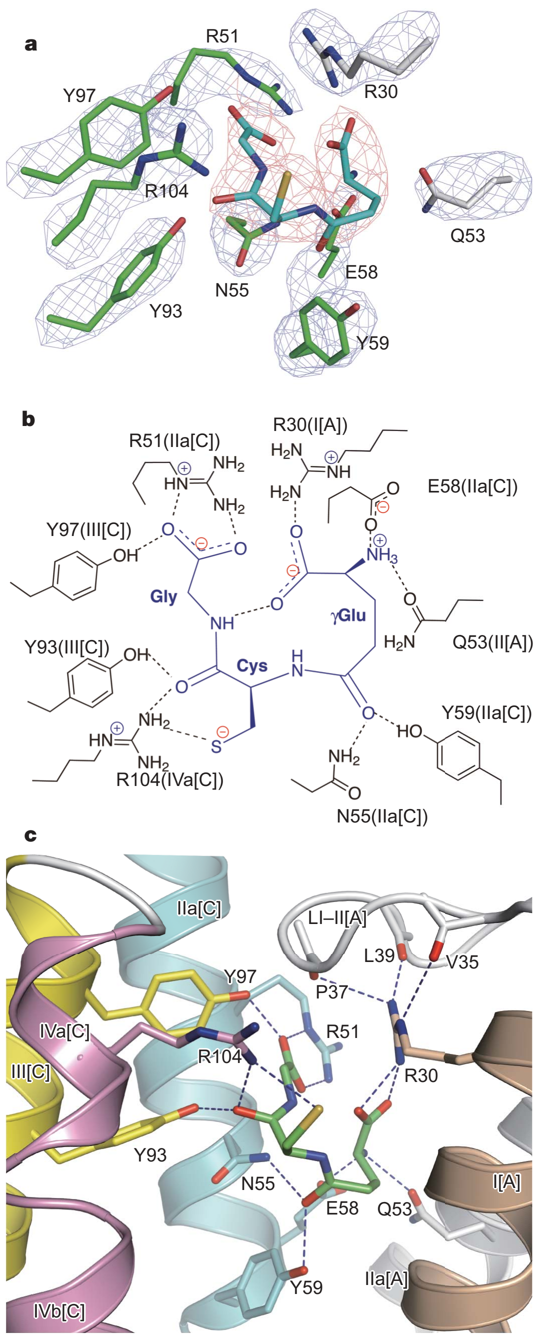


Figure 2 | GSH-binding site at the interface of monomers C and A. a, The polar side chains of LTC_4S interacting with bound GSH, with the $|F_o| - |F_c|$ omitted difference map (4σ). The electron density map for GSH (red) and the side chains (blue) is shown. Carbon atoms of the side chains of monomers C and A are shown in green and silver, respectively. **b**, **c**, Schematic drawing (**b**) and model (**c**) of electrostatic interactions of the side chains with a bound GSH. Nitrogen and oxygen atoms are shown in blue and red, respectively, in **a** and **c**. The GSH-binding site is composed of helices IIa (light blue), III (yellow) and IVa (pink) of monomer C, and helices I (light brown) and IIa (silver) of monomer A. Loop I–II of monomer A (silver) interacts with Arg 30 (R30).

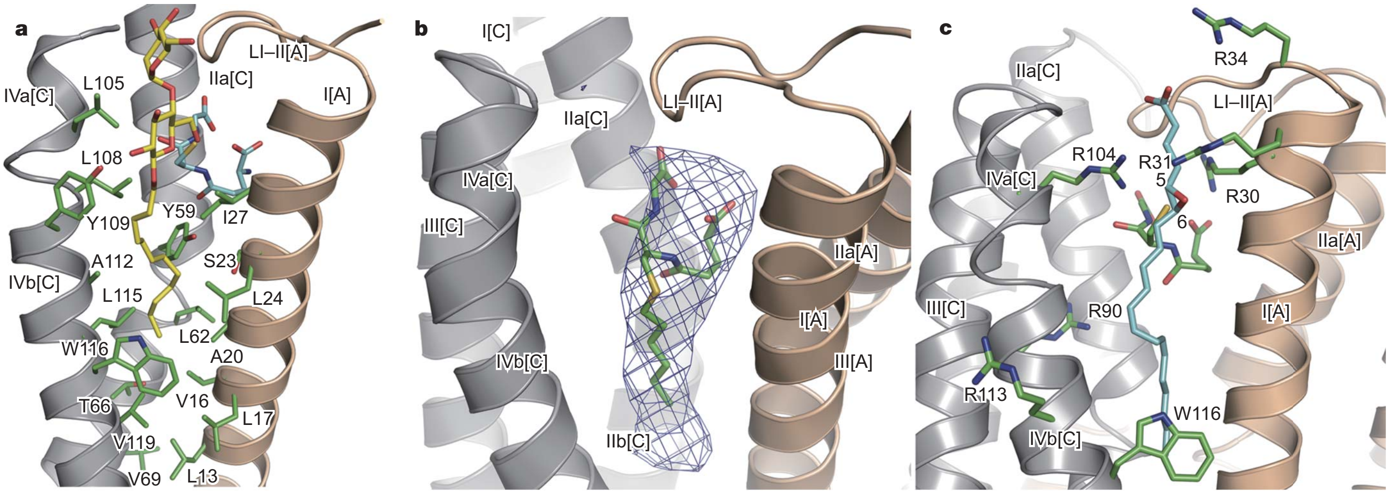


Figure 3 | The V-shaped valley as the putative LTA_4-binding site. a, A DDM molecule inserted in the V-shaped valley in LTC_4S with GSH. The bound DDM (yellow-coloured carbons) and GSH (light blue-coloured carbons) are represented in a stick model. Amino-acid residues (green carbons) forming the valley are from monomers C (silver) and A (light brown). **b**, Electron density map of bound *S*-hexyl GSH at 5.0 Å resolution. A difference Fourier map (2σ) with a predicted molecular conformation of *S*-hexyl GSH is shown. Monomers C and A are shown in silver and light brown, respectively. **c**, Docking model of LTA_4 into the binding site. The modelled LTA_4 (sky-blue carbons) was docked into the native LTC_4S structure on the basis of the bound DDM conformation. The extended conformation of LTA_4 fits well into the V-shaped valley. C6 and C5 of LTA_4 are placed in the vicinity of the thiolate anion (yellow) of GSH and the guanidino group of Arg 31 (R31), respectively.

LTC_4S, binds at each V-shaped valley between two neighbouring monomers in the trimer (Fig. 3a, and Supplementary Fig. 3). The V-shaped valley opens between the kinked helix IV of monomer C and helix I of monomer A at their interface, allowing the aliphatic chain of DDM to be inserted. The amino-acid residues Tyr 59, Leu 62, Thr 66 and Val 69 of helix II, and Leu 105, Leu 108, Tyr 109, Ala 112, Leu 115, Trp 116 and Val 119 of helix IV of monomer C, and Leu 13, Val 16, Leu 17, Ala 20, Ser 23, Leu 24 and Ile 27 of helix I of monomer A compose the hydrophobic interior of the valley. This hydrophobic milieu accommodates the hydrophobic aliphatic chain of DDM. At the bottom of the valley there is a hydrophobic space surrounded by Leu 62 and Val 69 of helix IIb and Trp 116 of helix IVb of monomer C, and by Leu 13, Val 16 and Leu 17 of helix I of monomer A. The side chain of Trp 116 of helix IV of monomer C separates the space from the lipid bilayer.

There is a tight hydrophobic interaction between the aliphatic chain from the end to the sixth carbon atom of the bound DDM and the residues at the bottom of the valley. The bound DDM protrudes from the valley and the middle glucoside moiety overhangs the bound GSH. The number of atoms from the tail end to the ring oxygen of the middle glucoside moiety is 15 in the DDM; this is equal to the number of atoms from the tail end to the carbon atom bonding to the epoxy oxygen in the substrate LTA_4. Thus, the bound structure of DDM to the V-shaped valley mimics a possible binding mode for LTA_4.

Recombinant LTC_4S was co-crystallized with *S*-hexyl GSH, a product mimic inhibitor for GSTs, and the structure was determined at 5.0 Å resolution by the molecular replacement method[27] with the LTC_4S structure (Fig. 3b). The *S*-hexyl GSH occupied the same site as GSH and its alkyl chain fitted well in the V-shaped active site of the native crystal. Thus, the binding mode of *S*-hexyl GSH in the LTC_4S complex is compatible with the DDM-binding mode in mimicry of the binding of LTA_4.

In the crystal LTC_4S structure, the LTA_4 binding model is constructed on the basis of the DDM conformation, with support from the binding mode of *S*-hexyl GSH (Fig. 3c). The aliphatic chain of the bound LTA_4 molecule is in an extended conformation with a bent structure in which two extended parts of the chain, C1 to C12 and C14 to C20, are connected by the two *cis* double bonds at C11 and C14 as a hinge. The tail of C14 to C20 of LTA_4 inserts into the hydrophobic space at the bottom of the V-shaped valley, reminiscent of the tail of the aliphatic chain of DDM observed in the crystal structure. In the LTA_4 docked model, the C6 bonding to the epoxy oxygen comes next to the thiol group of GSH, and the epoxy oxygen locates on the opposite side. The side chain of Arg 31 of monomer A locates near the epoxy oxygen of LTA_4.

On the basis of the LTA_4 docking model, we propose the general acid–base catalysis mechanism, in which Arg 31 as a general acid from helix I of monomer A and Arg 104 as a general base from helix III of monomer C catalyse the conjugation between the two substrates, GSH and LTA_4. The side chain of Arg 104 abstracts a proton to stabilize the formation of the thiolate anion of the bound GSH as the reactive species at physiological pH. The thiolate anion attacks the opposite side of C6 of LTA_4 from the epoxy bond to form a thioether bond in an S_N2 manner, with a chirality inversion at C6 from the *S*-configuration to the *R*-configuration. The protonation of the resultant hydroxyl anion at C5 by Arg 31 as the general acid at the other side in a concerted regiospecific and stereospecific reaction results in conjugation to form LTC_4.

We interpret the crystal structure of LTC_4S in a complex with GSH to reflect its constitutive state. The trimeric structure with binding of GSH molecules between adjacent monomers provides a hydrophobic environment in which LTA_4 can dock after its generation by the 5-lipoxygenase pathway of arachidonic acid metabolism as a result of cell activation. As the expression of LTC_4S is limited to haematopoietic cell types, whereas the receptors for LTC_4 and its metabolites are distributed to both the haematopoietic and stromal lineages with constitutive and induced responses, selective regulation of LTC_4S function could be therapeutic and would focus only on cells activated to provide LTA_4.

METHODS SUMMARY

Recombinant human LTC_4S with a His_6 tag at the C terminus was overexpressed in fission yeast as described[18]. After being solubilized with sodium deoxycholic acid and DDM, the enzyme was purified with four steps of column chromatography, including affinity chromatography with *S*-hexyl GSH. LTC_4S was crystallized in the presence of DDM and GSH with the sitting-drop vapour-diffusion technique at 4 °C for about 20 days. The grown crystals were $0.2 \times 0.2 \times 0.05\ mm^3$ in size. To prepare the SeMet LTC_4S, a Leu121Met mutation was created by site-directed mutagenesis. The Leu121Met yeast clone was grown with SeMet, and the SeMet-Leu121Met mutant LTC_4S was purified and crystallized similarly to the wild-type enzyme. Crystal structure of the human LTC_4S was determined by the multiwavelength anomalous diffraction (MAD) method.

The diffraction images from the SeMet derivative of Leu121Met were measured at 90 K at four X-ray wavelengths (0.96423 Å, 0.97904 Å, 0.97949 Å and 0.99519 Å). The electron density map showing a clear protein boundary was obtained successfully by the MAD phase calculation and after density modification at 3.3 Å resolution. The initial atomic model was constructed on the basis of the modified electron density map for the SeMet-Leu121Met mutant. The crystal structure of the native LTC_4S was deciphered by the molecular replacement method by AMoRe/CCP4 (ref. 27) using the initial atomic model and the iterative crystallographic structural refinement at 3.3 Å resolution by O[28], CNS[29] and REFMAC5/CCP4 (ref. 27). The crystallographic R and R_{free} values, and root mean squared deviations in lengths and angles in the current atomic model of LTC_4S are 0.229 and 0.257, and 0.003 Å and 0.568°, respectively. All amino-acid residues modelled exist in the favoured or allowed regions of the Ramachandran plot from PROCHECK/CCP4 (ref. 27).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 January; accepted 17 May 2007.
Published online 15 July 2007.

1. Kellaway, C. H. & Trethewie, W. R. The liberation of a slow reacting smooth muscle-stimulating substance in anaphylaxis. *Q. J. Exp. Physiol.* **30,** 121–145 (1940).
2. Brocklehurst, W. E. The release of histamine and formation of a slow-reacting substance (SRS-A) during anaphylactic shock. *J. Physiol. (Lond.)* **151,** 416–435 (1960).
3. Dahlen, S. E., Hedqvist, P., Hammarstrom, S. & Samuelsson, B. Leukotrienes are potent constrictors of human bronchi. *Nature* **288,** 484–486 (1980).
4. Dahlen, S. E. *et al.* Leukotrienes promote plasma leakage and leukocyte adhesion in postcapillary venules: *in vivo* effects with relevance to the acute inflammatory response. *Proc. Natl Acad. Sci. USA* **78,** 3887–3891 (1981).
5. Soter, N. A., Lewis, R. A., Corey, E. J. & Austen, K. F. Local effects of synthetic leukotrienes (LTC_4, LTD_4, LTE_4, and LTB_4) in human skin. *J. Invest. Dermatol.* **80,** 115–119 (1983).
6. Beller, T. C. *et al.* Cysteinyl leukotriene 1 receptor controls the severity of chronic pulmonary inflammation and fibrosis. *Proc. Natl Acad. Sci. USA* **101,** 3047–3052 (2004).
7. Kim, D. C. *et al.* Cysteinyl leukotrienes regulate Th2 cell-dependent pulmonary inflammation. *J. Immunol.* **176,** 4440–4448 (2006).
8. Weiss, J. W. *et al.* Bronchoconstrictor effects of leukotriene C in humans. *Science* **216,** 196–198 (1982).
9. Manning, P. J. *et al.* Inhibition of exercise-induced bronchoconstriction by MK-571, a potent leukotriene D_4-receptor antagonist. *N. Engl. J. Med.* **323,** 1736–1739 (1990).
10. Kanaoka, Y., Maekawa, A., Penrose, J. F., Austen, K. F. & Lam, B. K. Attenuated zymosan-induced peritoneal vascular permeability and IgE-dependent passive cutaneous anaphylaxis in mice lacking leukotriene C_4 synthase. *J. Biol. Chem.* **276,** 22608–22613 (2001).
11. Lam, B. K., Penrose, J. F., Freeman, G. J. & Austen, K. F. Expression cloning of a cDNA for human leukotriene C_4 synthase, an integral membrane protein conjugating reduced glutathione to leukotriene A_4. *Proc. Natl Acad. Sci. USA* **91,** 7663–7667 (1994).
12. Welsch, D. J. *et al.* Molecular cloning and expression of human leukotriene-C_4 synthase. *Proc. Natl Acad. Sci. USA* **91,** 9745–9749 (1994).
13. Bresell, A. *et al.* Bioinformatic and enzymatic characterization of the MAPEG superfamily. *FEBS J.* **272,** 1688–1703 (2005).
14. Samuelsson, B., Dahlen, S. E., Lindgren, J. A., Rouzer, C. A. & Serhan, C. N. Leukotrienes and lipoxins: structures, biosynthesis, and biological effects. *Science* **237,** 1171–1176 (1987).
15. Jakobsson, P. J., Mancini, J. A. & Ford-Hutchinson, A. W. Identification and characterization of a novel human microsomal glutathione S-transferase with leukotriene C_4 synthase activity and significant sequence identity to 5-lipoxygenase-activating protein and leukotriene C_4 synthase. *J. Biol. Chem.* **271,** 22203–22210 (1996).
16. Jakobsson, P. J., Mancini, J. A., Riendeau, D. & Ford-Hutchinson, A. W. Identification and characterization of a novel microsomal enzyme with glutathione-dependent transferase and peroxidase activities. *J. Biol. Chem.* **272,** 22934–22939 (1997).
17. Yoshimoto, T., Soberman, R. J., Spur, B. & Austen, K. F. Properties of highly purified leukotriene C_4 synthase of guinea pig lung. *J. Clin. Invest.* **81,** 866–871 (1988).
18. Schmidt-Krey, I. *et al.* Human leukotriene C_4 synthase at 4.5 Å resolution in projection. *Structure* **12,** 2009–2014 (2004).
19. Holm, P. J. *et al.* Structural basis for detoxification and oxidative stress protection in membranes. *J. Mol. Biol.* **360,** 934–945 (2006).
20. Thoren, S. *et al.* Human microsomal prostaglandin E synthase-1: purification, functional characterization, and projection structure determination. *J. Biol. Chem.* **278,** 22199–22209 (2003).
21. Christmas, P., Weber, B. M., McKee, M., Brown, D. & Soberman, R. J. Membrane localization and topology of leukotriene C_4 synthase. *J. Biol. Chem.* **277,** 28902–28908 (2002).
22. Armstrong, R. N. Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* **10,** 2–18 (1997).
23. Lam, B. K., Penrose, J. F., Xu, K., Baldasaro, M. H. & Austen, K. F. Site-directed mutagenesis of human leukotriene C_4 synthase. *J. Biol. Chem.* **272,** 13923–13928 (1997).
24. Guillen Schlippe, Y. V. & Hedstrom, L. A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* **433,** 266–278 (2005).
25. Zimmer, J. S., Dyckes, D. F., Bernlohr, D. A. & Murphy, R. C. Fatty acid binding proteins stabilize leukotriene A_4: competition with arachidonic acid but not other lipoxygenase products. *J. Lipid Res.* **45,** 2138–2144 (2004).
26. Marfat, A. & Corey, E. J. Synthesis and structure elucidation of leukotrienes. *Adv. Prostaglandin Thromboxane Leukot. Res.* **14,** 155–228 (1985).
27. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50,** 760–763 (1994).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
29. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Tsurumura and N. Takahashi for their technical assistance. This work was supported by Tanpaku3000 of MEXT (to M.M.), RIKEN Matching fund (to H.A. and M.M.) and National Institutes of Health grants (to Y.K., B.K.L. and K.F.A.).

Author Contributions H.A., Y.K. and D.I. contributed equally to this work. Y.K. overexpressed and purified the human LTC_4S. H.A. optimized the enzyme purification methods. H.A., D.I. and T.S. crystallized the enzyme. D.I. prepared the SeMet protein. H.A. and M.M. solved the structure. B.K.L. performed the LTC_4S assay. H.A., Y.K., K.F.A. and M.M. guided the work and prepared the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank with the accession number 2PNO. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (miyano@spring8.or.jp) or K.F.A. (fausten@rics.bwh.harvard.edu).

METHODS

Purification and crystallization of human LTC_4S. Recombinant human LTC_4S with a His_6 tag at the C terminus was overexpressed in fission yeast as described[18] and purified with four steps of column chromatography for the following crystallization. The yeast pellets were disrupted on ice by glass beads in the lysis buffer (50 mM HEPES-NaOH pH 8.0, 1 mM EDTA, 10 mM 2-mercaptoethanol) and centrifuged at 6,000*g* for 15 min at 4 °C. The supernatant of the cell lysate was adjusted to 10% (v/v) glycerol, 0.5% (w/v) sodium deoxycholic acid and 1% (w/v) DDM (Anatrace). The mixture was stirred gently for 1 h for solubilization of the LTC_4S, and then centrifuged at 17,300*g* for 20 min at 4 °C.

First, *S*-hexyl GSH-agarose (Sigma) was used for affinity chromatography. The supernatant was incubated with the agarose gel for 1 h at 4 °C with gentle rotation to adsorb the solubilized LTC_4S. The gel was then washed seven times to remove sodium deoxycholic acid, decrease DDM stepwise and equilibrate with 20 mM Tris-HCl pH 7.5, 0.04% (w/v) DDM, 1 mM 2-mercaptoethanol and 10% (v/v) glycerol (buffer A). The gel was further washed with buffer A containing 5 mM GSH, and the bound LTC_4S was eluted with buffer A containing 5 mM GSH and 30 mM probenecid.

Second, Ni-chelating resin (Qiagen) was used to remove probenecid. The eluate was incubated with the Ni-chelating resin for 1 h at 4 °C with gentle stirring. The resin was washed with 20 mM MES-NaOH pH 6.5, 0.04% (w/v) DDM, 1 mM 2-mercaptoethanol, 10% (v/v) glycerol and 5 mM GSH (buffer B) until the ultraviolet absorbance of probenecid at 280 nm returned to the baseline in the eluted buffer B. Bound LTC_4S was eluted with buffer B containing 250 mM imidazole.

Third, the eluate was subjected to Superose12 (GE Healthcare Bio-Science) gel filtration chromatography equilibrated in buffer B containing 0.1 M NaCl. LTC_4S was eluted as a symmetrical single peak at the elution volume corresponding to 63 kDa as apparent molecular mass. Fourth, the eluted fraction was applied to a PD10 column (GE Healthcare Bio-Science) equilibrated in sample buffer (20 mM MES-NaOH pH 6.5, 0.04% (w/v) DDM, 5 mM GSH). LTC_4S was further concentrated to 6.5 mg ml^{-1} by using an ultrafiltration membrane with a 50-kDa cutoff for the crystallization.

LTC_4S was crystallized in an equi-volume mixture of 6.5 mg ml^{-1} LTC_4S in 20 mM MES pH 6.5, 0.04% (w/v) DDM and 5 mM GSH and reservoir solution (0.1 M HEPES pH 7.5, 30% (v/v) PEG400) with the sitting-drop vapour-diffusion technique at 4 °C for about 20 days. The grown crystals with a size of $0.2 \times 0.2 \times 0.05$ mm^3 were mounted on a cryo-loop and frozen with liquid nitrogen.

Preparation of Leu121Met mutant and its selenomethionine derivative. To prepare SeMet LTC_4S, a Leu121Met mutation was created with the use of a QuikChange site-directed mutagenesis kit (Stratagene) with sense (5′-TGGCTGCTGGTGGCGATGGCTGCGCTCGGCCTG-3′) and antisense (5′-CAGGCCGAGCGCAGCCATCGCCACCAGCAGCCA-3′) primers. The resultant plasmid was transfected into *Schizosaccharomyces pombe*, h^- *leu1-32*, and stable clones were established.

The Leu121Met mutant clone was grown overnight in SC medium (Q-BIOgene) supplemented with glucose and yeast nitrogen base (Difco). The yeast was then cultured in EMM medium (Q-BIOgene) containing 37.3 mg l^{-1} SeMet at 30 °C for 20 h. The SeMet-Leu121Met mutant was purified and crystallized as described for native LTC_4S except that the protein was equilibrated in a sample buffer, 20 mM MES-NaOH pH 6.5, 0.04% (w/v) DDM, 5 mM GSH, 3.3% (v/v) glycerol, 33 mM NaCl.

The final protein concentration was 6.6 mg ml^{-1}. The size of the harvested crystals was $0.2 \times 0.2 \times 0.05$ mm^3.

Structure determination. The crystal structure of the human LTC_4S was determined by the MAD method. The diffraction images from the SeMet derivative of Leu121Met were measured at BL44B2/SPring-8 (ref. 30) at 90 K at four X-ray wavelengths (0.96423 Å, 0.97904 Å, 0.97949 Å and 0.99519 Å). The collected images were reduced by MOSFLM/CCP4 (ref. 27) and SCALA/CCP4 (ref. 27). The electron density map showing a clear protein boundary was obtained successfully by the MAD phase calculation and after density modification at 3.3 Å resolution with SHELXE[31]. The further electron density modification including the electron density averaging between the 12 monomers was done by DM/CCP4 (ref. 27). The initial atomic model was constructed on the basis of the modified electron density map for the SeMet-Leu121Met mutant. The crystal structure of the native LTC_4S was deciphered by the molecular replacement method by AMoRe/CCP4 (ref. 27) using the initial atomic model and the iterative crystallographic structural refinement at 3.3 Å resolution by O[28], CNS[29] and REFMAC5/CCP4 (ref. 27). The crystallographic R and R_{free} values, and root mean squared deviations in lengths and angles in the current atomic model of LTC_4S are 0.229 and 0.257, and 0.003 Å and 0.568°, respectively. All amino-acid residues modelled exist in the favoured or allowed regions of the Ramachandran plot from PROCHECK/CCP4 (ref. 27).

30. Adachi, S. *et al.* The RIKEN structural biology beamline II (BL44B2) at the SPring-8. *Nucl. Instrum. Methods Phys. Res. A* **467–468,** 711–714 (2001).
31. Sheldrick, G. M. Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217,** 644–650 (2002).

LETTERS

Structural basis for synthesis of inflammatory mediators by human leukotriene C_4 synthase

Daniel Martinez Molina[1,4], Anders Wetterholm[2], Andreas Kohl[1], Andrew A. McCarthy[5], Damian Niegowski[1,4], Eva Ohlson[2], Tove Hammarberg[2], Said Eshaghi[1], Jesper Z. Haeggström[2] & Pär Nordlund[1,3]

Cysteinyl leukotrienes are key mediators in inflammation and have an important role in acute and chronic inflammatory diseases of the cardiovascular and respiratory systems, in particular bronchial asthma. In the biosynthesis of cysteinyl leukotrienes, conversion of arachidonic acid forms the unstable epoxide leukotriene A_4 (LTA_4). This intermediate is conjugated with glutathione (GSH) to produce leukotriene C_4 (LTC_4) in a reaction catalysed by LTC_4 synthase[1]: this reaction is the key step in cysteinyl leukotriene formation. Here we present the crystal structure of the human LTC_4 synthase in its apo and GSH-complexed forms to 2.00 and 2.15 Å resolution, respectively. The structure reveals a homotrimer, where each monomer is composed of four transmembrane segments. The structure of the enzyme in complex with substrate reveals that the active site enforces a horseshoe-shaped conformation on GSH, and effectively positions the thiol group for activation by a nearby arginine at the membrane–enzyme interface. In addition, the structure provides a model for how the ω-end of the lipophilic co-substrate is pinned at one end of a hydrophobic cleft, providing a molecular 'ruler' to align the reactive epoxide at the thiol of glutathione. This provides new structural insights into the mechanism of LTC_4 formation, and also suggests that the observed binding and activation of GSH might be common for a family of homologous proteins important for inflammatory and detoxification responses.

Leukotriene (LT) C_4 synthase is the pivotal enzyme in the biosynthesis of the classical 'slow reacting substance of anaphylaxis', which comprises LTC_4, LTD_4 and LTE_4 (ref. 2). These molecules are powerful lipid mediators of inflammation and immediate hypersensitivity reactions[3,4]. Typically, cysteinyl leukotrienes are produced by activated mast cells, eosinophils and monocytes and elicit smooth muscle contractions at nanomolar concentrations in the respiratory tract and microcirculation[5–8]. Today, they are well known bronchoconstrictive and oedema promoting agents and indeed, drugs targeting cysteinyl leukotriene receptors have offered therapeutic opportunities in the medical treatment of asthma[9]. However, all patients do not respond to receptor antagonist treatment[10]. Therefore there is a great need for more selective therapeutic agents.

We have determined the crystal structures of human LTC_4 synthase in its apo and GSH-complexed forms. LTC_4 synthase is an 18 kDa integral membrane enzyme, which, together with the highly homologous 5-LO activating protein (FLAP), the more distantly related microsomal glutathione-*S*-transferases (MGSTs) and microsomal prostaglandin E_2 synthase type 1 (MPGES-1), make up the MAPEG (membrane-associated proteins in eicosanoid and glutathione metabolism) protein superfamily[11]. The structure of LTC_4 synthase reveals a compact trimeric protein, where each subunit is composed of five long α-helices. The first four helices form the transmembrane segments, while helix 5 extends out of the membrane plane (Fig. 1a). A long loop (loop 1) connecting helices 1 and 2 folds on top of the neighbouring monomer, contributing to subunit interactions (Fig. 1b).

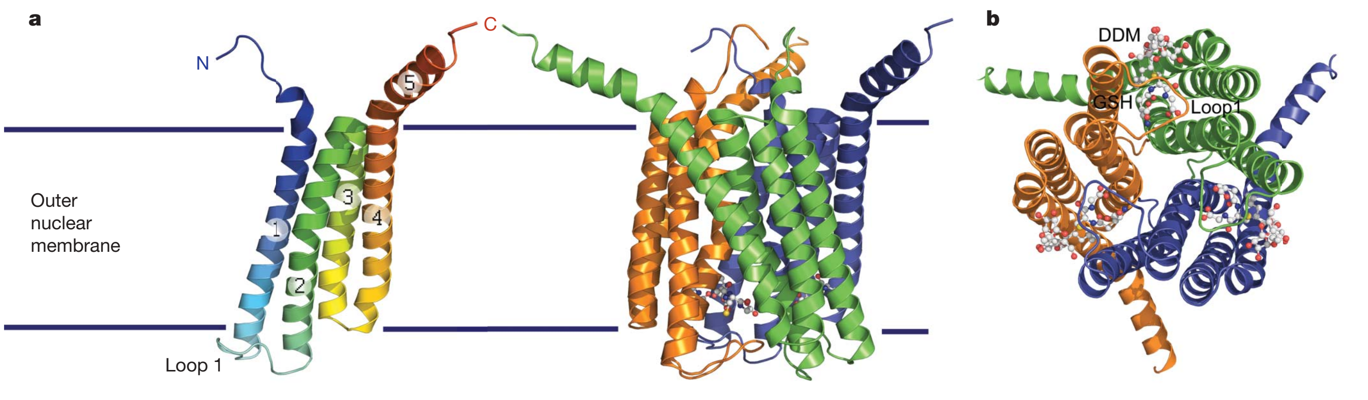


Figure 1 | LTC_4 synthase overall structure. **a**, Left, front view of LTC_4 synthase protomer in ribbon representation showing helices 1–5 in successive colouring. The entire LTC_4 synthase polypeptide, except for the last residue (the last four in the GSH-complexed structure), can be traced in the electron density maps. Right, the homotrimer with protomers coloured orange, green and blue. Bound glutathione, indicating the position of the active site, is in ball and stick representation. Approximate membrane positions indicated by blue lines. **b**, Cytosolic view of the trimer, showing the position of the three active sites with loop 1 covering each binding pocket. Glutathione is shown in ball and stick representation together with modelled detergent (DDM).

[1]Division of Biophysics, [2]Division of Chemistry 2, Department of Medical Biochemistry and Biophysics, [3]Structural Genomics Consortium, Karolinska Institutet, 17177 Stockholm, Sweden. [4]Department of Biochemistry and Biophysics, Stockholm University, S-106 91 Stockholm, Sweden. [5]European Molecular Biology Laboratory, Grenoble outstation, 6 Rue Jules Horowitz, 38042 Grenoble Cedex 9, France.

The amino-terminal 6-His tag used for purification is clearly visible in the structure, and forms a unique metal coordinated cluster, composed of 12 crystallographically related hexa-histidine tags that coordinate 8 metals (Supplementary Fig. 2a, b).

A number of extended densities, likely to originate from ordered detergents or lipids, are seen in the electron density maps and have been introduced into the crystallographic refinement as the detergent dodecyl maltoside (DDM) used in the purification, or as tentative aliphatic chains. A notable cluster of such aliphatic chains pack against helices 1, 3 and 5, underscoring the high hydrophobicity of helix 5, where the first half is almost exclusively composed of hydrophobic residues (Fig. 2a). In its natural membrane environment, it is plausible that helix 5 slightly changes direction to be embedded into the lipid bilayer as an interfacial surface helix.

The active site of LTC_4 synthase, identified by a bound GSH, is buried at the interface of two adjacent monomers close to the membrane face where loop 1 is positioned. It is likely that this part of the protein faces the cytoplasmic side of the outer nuclear membrane. This would facilitate delivery and release of substrates and product, as both the preceding and following steps in the synthesis pathway are conducted on the cytosolic side of the membrane.

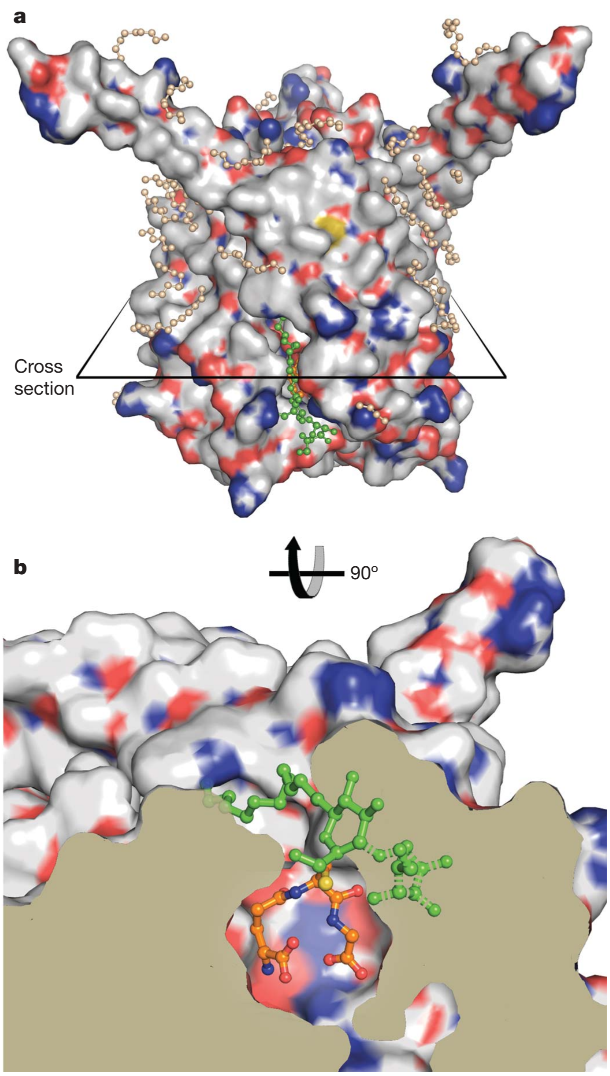


Figure 2 | Substrate–protein and lipid–protein interactions. a, Surface representation of the trimeric protein coloured by molecule. In light brown are ball and stick representations of carbon chains lining the protein surface. The detergent molecule is visualized in green and underneath is a glimpse of the bound glutathione in orange. **b,** Cross-section from the cytosolic side of LTC_4 synthase, as indicated in **a**, revealing the polar binding pocket of GSH and the cleft where the aliphatic co-substrate binds. Dashed bonds highlight the partial occupancy of the second pyranoside of the detergent molecule.

GSH is bound deep in a polar pocket at the interface between helix 1 and 2 from one monomer and 3 and 4 from a neighbouring monomer (residues from the latter are indicated with ′ (prime)) (Figs 1b, 2b). The GSH molecule makes polar interactions to residues from both monomers constituting the active site (Fig. 3a, b). The carboxylate moieties of GSH make salt bridges to Arg 51′ and Arg 30 at the base of the binding pocket, effectively bending GSH and directing its thiol group towards the membrane interface where it interacts with Arg 104′ and is positioned close to a bound DDM molecule (Figs 2b, 3a). Additional polar interactions to GSH are made by Gln 53, Asn 55′, Glu 58′, Tyr 59′, Tyr 93′ and Tyr 97′. Several non-polar interactions are also made (Ile 27, Pro 37 and Leu 108′), providing a optimal fit for GSH into its binding pocket (for a detailed overview of the GSH binding, see Supplementary Fig. 4). In the apo form, a tentative sulphate ion is located in this pocket. Comparison of the GSH-bound LTC_4 synthase structure with the apo-LTC_4 synthase structure reveals that only local adjustments of polar amino-acid side chains are made upon GSH binding (Fig. 3b). Loop 1 appears to partially cover the access of GSH to its binding pocket (Figs 1b, 3b). Therefore some flexibility of loop 1 might be required during the reaction cycle, consistent with structural rearrangement of this loop upon GSH binding (Fig. 3b). Crystallographic studies of LTC_4 synthase with the co-substrate LTA_4 is challenging, owing to the short half-life (~10 s) of this epoxide containing compound. However, we propose that the tentative DDM molecule bound close to GSH serves as a good substrate model, as it has important structural similarities to LTA_4. The proposed binding site provides a well-suited hydrophobic crevice for the lipophilic substrate where Trp 116′ forms a lid over the ω-end of the substrate and plays a key role in positioning the aliphatic chain. Together with residues from both subunits it seals off the crevice (Fig. 4a, b) and serves as a ruler, allowing the proper positioning of C6 of the LTA_4 epoxide near the reactive GSH thiol (Fig. 4a). A DDM molecule is also bound in a similar position in the apo structure but the ω-terminus of the detergent sits outside, rather than wedged into the Trp 116′ pocket (Fig. 4b). In addition, the detergent carbon atom corresponding to C6 of LTA_4 is positioned more than 8 Å away from where the GSH thiol would be located. This suggests that GSH binding assists in the formation of an appropriate lipid binding crevice for LTA_4, presumably by covering charged groups in the active site and extending the interaction surface for the lipid.

In order to promote the GSH conjugation reaction, the nucleophilicity of the GSH thiol is most probably enhanced by the enzyme. Arg 104 is ideally positioned to promote the pK_a shift of the GSH thiol through its positive charge, where one of the η-nitrogens is able to mediate a polar interaction with the GSH sulphur (3.2 Å). This unusually short sulphur–nitrogen distance, together with the lack of additional hydrogen bond acceptors, suggest that the enzyme-bound thiol group may in fact be an anionic thiolate in the crystal structure, as suggested for the distantly related MGST-1 (ref. 12). The GSH thiol is well positioned for a nucleophilic attack on the allylic C6 of the oxirane ring of LTA_4 (ref. 13). There are no well positioned acid residues situated in the active site that could provide a proton to the evolving C–O group of the substrate. It is therefore likely that the substrate oxyanion is stabilized by hydrogen bonding with the enzyme. Arg 104 is located close to the expected position of the C5 of the substrate and could therefore also assist in stabilizing the substrate anion. In addition, Arg 31 is located on the distal side of the substrate and although it is flexible in the present structure, this residue may also assist in the stabilization of the substrate anion. The regiospecificity of the resulting S_N2 mechanism will be defined by the productive binding of the epoxide in the active site. In the reaction, the opening of the epoxide will result in chiral inversion such that the 6*S* stereochemistry of the epoxide oxygen of LTA_4 will

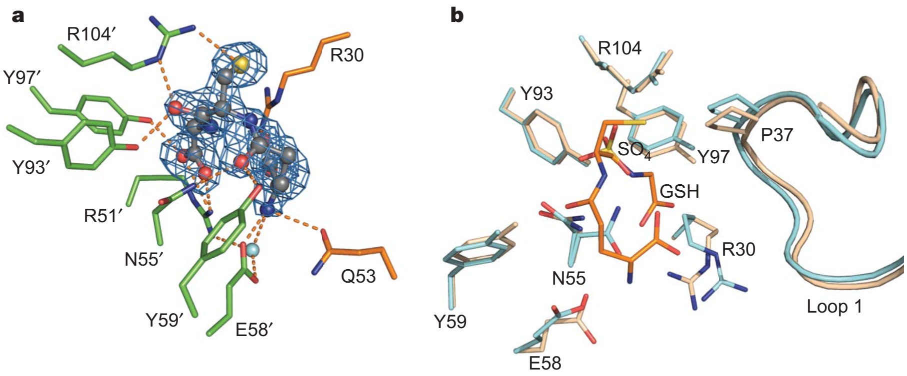


Figure 3 | Glutathione binding. **a**, Electron density map in blue ($F_o - F_c$, contoured at 3σ and phased with the apo structure before refinement) for bound glutathione shown in ball and stick representation. Interacting side chains are coloured by monomer and labelled accordingly. Chemical bonds to glutathione are drawn as dashed lines. A coordinated water molecule is shown as a light blue sphere. **b**, Superposition of the active sites in the apo and glutathione-bound structures in blue and light brown, respectively. Glutathione is shown in orange stick representation and the sulphate molecule from the apo structure in yellow.

be completely converted to the 6*R* configuration of the resulting glutathionyl moiety of the product LTC_4.

Figure 4c depicts a schematic summary of the proposed mechanism for substrate binding and product generation for LTC_4 synthase. In the proposed binding scheme, GSH enters its binding pocket from the cytosol, and by doing so enables binding of LTA_4 that previously resided in the lipid membrane. The hydrophilic addition to LTA_4 enables LTC_4 to migrate out in the cytosol, perhaps depending on the flexibility of loop 1.

The structural information on LTC_4 synthase, especially on the location of the transmembrane segments and the position of the active site residues, allows for a structure based alignment of the MAPEG family (Supplementary Fig. 5). Previous sequence alignments suggest a common evolutionary origin of the human MAPEG members, where sequence conservation is low but significant in the two central transmembrane segments, helix 2 and helix 3 (ref. 14). Sequences in the terminal transmembrane segments, helix 1 and 4, however, are more diverged and clear similarities can only be seen within subfamilies. In the structure based alignment, further conservation is added to the active site region of the MAPEG family by the likely alignment of helix 1. The alignment of the first three helices strongly supports the idea that the majority of the GSH coordinating residues are conserved in all the GSH dependent MAPEG members (all MAPEG members except FLAP are known to catalyse GSH dependent reactions). Out of the 8 polar residues coordinating GSH from these helices, 5 are fully conserved in the GSH dependent enzymes, while the remaining 3 have mainly conservative amino-acid substitutions. Our alignment of helix 4 is more tentative. Arg 104, the critical catalytic residue proposed here, is aligned with arginines predicted to reside in the N-terminal end of helix 4 of other GSH dependent MAPEG members. The position of this potential catalytic Arg in MGST-1 (Arg 130) appears to be consistent with a low-resolution structure of MGST-1 (ref. 14). Whereas the amino acids

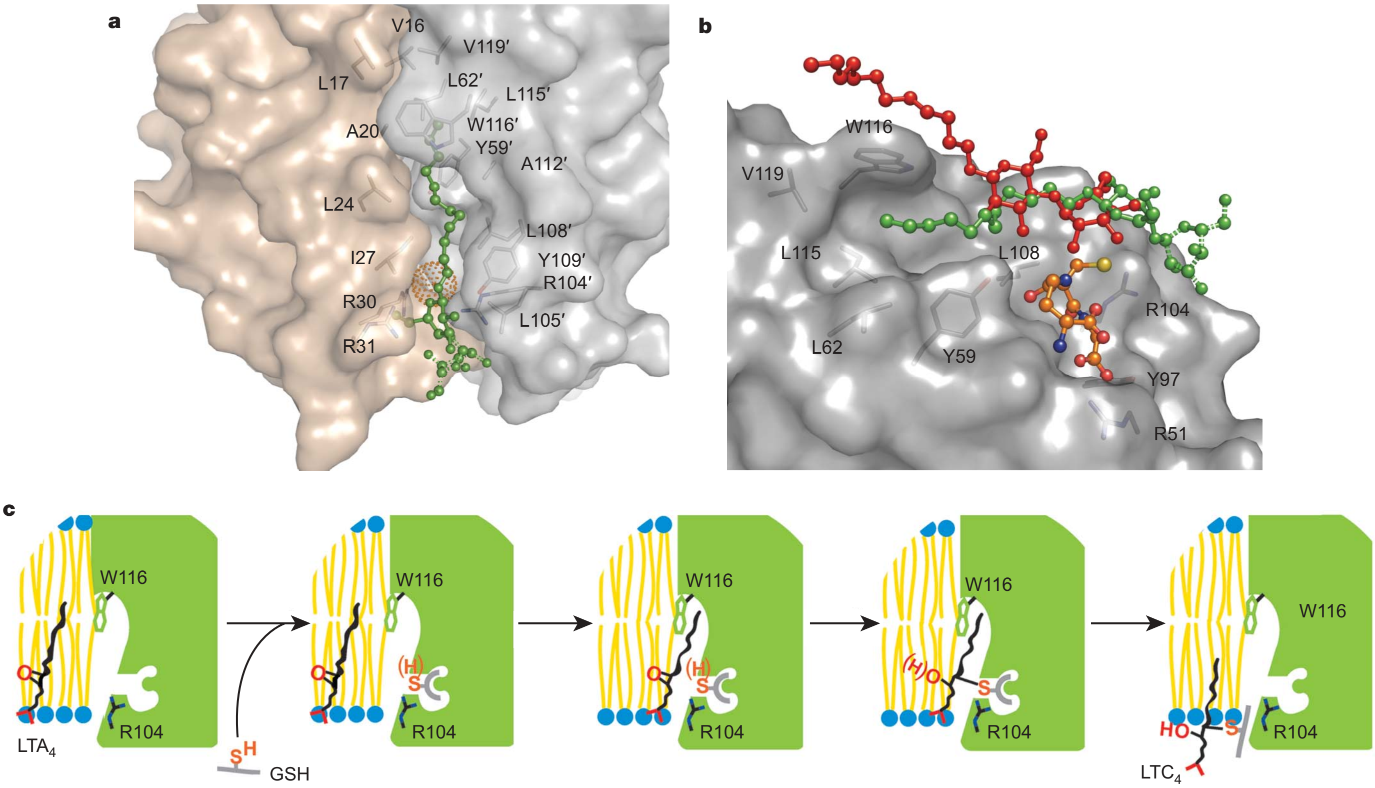


Figure 4 | The hydrophobic binding crevice and proposed schematic mechanism of substrate binding and catalysis. **a**, The interface between two monomers shown in transparent grey and brown with interacting and bordering residues shown as sticks. An orange sphere shows the area where the cysteinyl sulphur is located. The bound detergent is depicted as a green ball and stick model. **b**, Side view of the substrate binding region rotated 90° anticlockwise with respect to **a**. For clarity, only the grey monomer, harbouring the Trp116′ crevice, is depicted. The removed monomer contributes hydrophobic residues to the lipid binding cleft and polar residues to glutathione. Glutathione backbone is coloured orange, the detergent is green, as in **a**, for the glutathione bound structure and red for the apo structure. **c**, Schematic view of the proposed substrate binding and conjugation. H in brackets shows that a thiolate might be generated already upon binding of GSH and that the initial product might be anionic.

coordinating GSH are highly conserved in the MAPEG family, the residues lining the proposed LTA_4 binding crevice are not. Although most of the corresponding residues are hydrophobic in the other families, their size and aromatic composition vary.

This may indicate how the MAPEG members have evolved to recognize different lipid substrates binding in the same pocket. However, it could also suggest that lipid recognition occurs differently in the other family members: for example, MGST-1 has evolved to be highly unspecific. In addition, the nature of the head group of LTA_4 has been shown to be variable without loss of enzymatic activity[15,16]. This lack of major specificity for the head group can presumably be explained by the lack of a distinct binding pocket for the carboxylate (Figs 2a, 4a).

In conclusion, it appears that the horseshoe-shaped binding mode of GSH is conserved throughout the GSH dependent MAPEG family members and, potentially, that the structural basis for GSH activation may be related. In LTC_4 synthase, the arginine at position 104 is likely to play a key role in activating the GSH thiol for a nucleophilic attack on the LTA_4 epoxide. The intriguing lipid binding pocket of LTC_4 synthase allows LTA_4 to fit into the active site by the Trp 116 crevice, effectively positioning the C6 of the substrate for conjugation with GSH. It is plausible that this principle of lipid substrate recognition is primarily based on the aliphatic chain length rather than the nature of the head group. This may have some bearing also for substrate recognition in other integral membrane enzymes.

METHODS SUMMARY

Human LTC_4 synthase was expressed as a hexa-histidine construct in the yeast *Pichia pastoris*. The protein was extracted with a Triton X-100/Triton-DOC mixture and purified using two affinity chromatography steps and gel filtration. Crystals were grown using the sitting drop vapour diffusion technique. GSH derivatives were obtained by mixing equal volumes of 6 mM GSH, dissolved in mother liquor, with protein and left to soak for 24 h at 4 °C.

All X-ray data were collected on beam lines ID14-4 and ID23-2 at the European Synchrotron Radiation Facility (ESRF). Diffraction data of native and GSH soaked crystals were processed and scaled using Mosflm[17] and SCALA[18] while the XDS[19] suite was used for heavy atom MAD (multiwavelength anomalous diffraction) data sets. Three highly redundant data sets to 3.2 Å were collected around the Pt LIII edge. A single platinum site was located in the peak data set using SHELXD after local scaling in XPREP and SHELXE[20] had confirmed the correct hand. Experimental MAD phases were obtained using the program SHARP[21]. These phases were further improved using a 75% solvent content in SOLOMON[22], resulting in an FOM (figure of merit) of 0.88. The initial map was used to model α-helices, which were in turn used together with high-resolution data to assign residues with ARP/wARP[23]. The initial model was further built using Coot[24] and used for molecular replacement with Phaser[25] of a higher resolution apo structure and the GSH bound structure. Both the apo and GSH bound structure were built in Coot and refined using Refmac[26] to an R_{work}/R_{free} of 17.6/21.2% and 18.3/22.0%, respectively. All images were prepared using Pymol[27] except Supplementary Fig. 4, which was generated with LIGPLOT[28]. Crystallographic data are provided in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 February; accepted 11 June 2007.
Published online 15 July 2007.

1. Lam, B. K. & Austen, K. F. Leukotriene C-4 synthase: A pivotal enzyme in cellular biosynthesis of the cysteinyl leukotrienes. *Prostaglandins Other Lipid Mediat.* **68-69,** 511–520 (2002).
2. Samuelsson, B. Leukotrienes: mediators of immediate hypersensitivity reactions and inflammation. *Science* **220,** 568–575 (1983).
3. Funk, C. D. Prostaglandins and leukotrienes: Advances in eicosanoid biology. *Science* **294,** 1871–1875 (2001).
4. Samuelsson, B., Dahlén, S.-E., Lindgren, J.-Å., Rouzer, C. A. & Serhan, C. N. Leukotrienes and lipoxins: structures, biosynthesis, and biological effects. *Science* **237,** 1171–1176 (1987).
5. Hanna, C. J., Bach, M. K., Pare, P. D. & Schellenberg, R. R. Slow-reacting substances (leukotrienes) contract human airway and pulmonary vascular smooth-muscle in vitro. *Nature* **290,** 343–344 (1981).
6. Dahlén, S.-E., Hedqvist, P., Hammarström, S. & Samuelsson, B. Leukotrienes are potent constrictors of human bronchi. *Nature* **288,** 484–486 (1980).
7. Dahlén, S.-E. *et al.* Leukotrienes promote plasma leakage and leukocyte adhesion in postcapillary venules: *In vivo* effects with relevance to the acute inflammatory response. *Proc. Natl Acad. Sci. USA* **78,** 3887–3891 (1981).
8. Weiss, J. W. *et al.* Bronchoconstrictor effects of leukotriene-C in humans. *Science* **216,** 196–198 (1982).
9. Drazen, J. F., Israel, E. & O'Byrne, P. Treatment of asthma with drugs modifying the leukotriene pathway. *N. Engl. J. Med.* **340,** 197–206 (1999).
10. Bisgaard, H. Leukotriene modifiers in pediatric asthma management. *Pediatrics* **107,** 381–390 (2001).
11. Jakobsson, P. J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A. & Persson, B. Common structural features of MAPEG — a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Sci.* **8,** 689–692 (1999).
12. Morgenstern, R., Svensson, R., Bernat, B. A. & Armstrong, R. N. Kinetic analysis of the slow ionization of glutathione by microsomal glutathione transferase MGST1. *Biochemistry (Mosc.)* **40,** 3378–3384 (2001).
13. Corey, E. J. *et al.* Stereospecific total synthesis of a ''slow reacting substance'' of anaphylaxis, leukotriene C4. *J. Am. Chem. Soc.* **102,** 1436–1439 (1980).
14. Holm, P. J. *et al.* Structural basis for detoxification and oxidative stress protection in membranes. *J. Mol. Biol.* **360,** 934–945 (2006).
15. Izumi, T. *et al.* Solubilization and partial purification of leukotriene C4 synthase from guinea-pig lung: A microsomal enzyme with high specificity towards 5,6-epoxide leukotriene A4. *Biochim. Biophys. Acta* **959,** 305–315 (1988).
16. Yoshimoto, T., Soberman, R. J., Spur, B. & Austen, K. F. Properties of highly purified leukotriene C4 synthase of guinea pig lung. *J. Clin. Invest.* **81,** 866–871 (1988).
17. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* **26,** 27 (1992).
18. Collaborative. Computational Project, Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
19. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
20. Uson, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9,** 643–648 (1999).
21. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494 (1997).
22. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
23. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
24. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
25. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61,** 458–464 (2005).
26. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
27. DeLano, W. L. The PyMOL Molecular Graphics System. (DeLano Scientific, Palo Alto, 2002).
28. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. Ligplot — a program to generate schematic diagrams of protein ligand interactions. *Protein Eng.* **8,** 127–134 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Swedish Research Council, the Swedish Cancer Society, the Göran Gustafsson foundation, the Knut och Alice Wallenberg foundation, the EU integrated projects E-MEP and EICOSANOX. This Letter reflects only the authors' views and the European Commission is not liable for any use that may be made of the information herein. We also thank staff at ESRF, BESSY, SLS and MAXLAB for beam time and assistance during data collection. We are grateful to T. Cornvik for help with artwork and to T. Bergman for help with mass spectrometry analysis.

Author Contributions A.W., T.H. and J.Z.H. cloned, expressed and purified LTC_4 synthase. Further purification and characterization of LTC_4 synthase in various detergents was performed by A.W., D.M.M. and E.O. and was supervised by S.E. D.M.M. crystallized LTC_4 synthase and solved the structure together with A.A.M. and A.K. Crystallographic work was performed by D.M.M. under supervision of A.K. and P.N. Extensive crystal and derivative screening was conducted by D.M.M., A.K. and D.N. Overall project management and strategy was performed by J.Z.H. and P.N. Writing of the manuscript was performed jointly by P.N., J.Z.H., S.E. and D.M.M. All authors discussed the results and commented on the manuscript.

Author Information Coordinates and structure factors for LTC_4 synthase in its apo form and in complex with GSH have been deposited in the Protein Data Bank with accession codes 2UUI and 2UUH, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.Z.H. (jesper.haeggstrom@ki.se), S.E. (said.eshaghi@ki.se) and P.N. (par.nordlund@ki.se).

METHODS

Materials. Imidazole, Tris base, NaCl, KCl, Triton X-100, sodium deoxycholate, S-hexylglutathione agarose, probenecid, reduced glutathione (GSH), and 2-mercaptoethanol were obtained from Sigma. Dodecyl maltoside was obtained from Anatrace.

Cloning and plasmid construction. The human *LTC4S* cDNA (IMAGE cDNA clone 5277851, MRC geneservice) was subcloned into pPICZA (Invitrogen). Both the cDNA, supplemented with an N-terminal sequence encoding a His_6 tag, and the vector were PCR amplified and the products were co-transformed into $CaCl_2$-competent *Escherichia coli* (TOP10, Invitrogen), using the endogenous recombinase activity of *E. coli* to recombine the fragments. The protein coding part of the resulting plasmid, pPICZ-hisLTC4S, was verified by DNA sequencing.

Protein expression and purification. The expression vector was transformed into *P. pastoris* KM71H cells using the Pichia EasyComp Transformation kit (Invitrogen). Recombinant cells were cultivated in baffled flasks in 2.5 l minimal yeast medium with glycerol (Invitrogen) at 27 °C. When OD_{600} reached 8–10, the cells were resuspended in 0.5 l minimal yeast medium with 0.5% methanol. The cells were harvested after 72 h by centrifugation (2,500*g*, 7 min) and resuspended in 50 mM Tris-HCl, pH 7.8, 100 mM KCl and 10% glycerol. The cells were homogenized with glass beads (0.5 mm) and the slurry was filtered through nylon net filters (180 μm, Millipore) and centrifuged (1,500*g*, 10 min). Membrane bound proteins in the supernatant were solubilized with Triton X-100 (1%, v/v) and sodium deoxycholate (0.5%, w/v) for 1 h with stirring on ice. After centrifugation (10,000*g*, 10 min) the supernatant was supplemented with 10 mM imidazole and loaded on Ni-Sepharose Fast Flow (GE Healthcare). The column was washed with buffer A (25 mM Tris-HCl, pH 7.8, 10% glycerol, 0.1% Triton X-100 and 5 mM 2-mercaptoethanol) supplemented with 20 mM imidazole and 0.1 M NaCl, followed by additional wash with buffer A containing 40 mM imidazole and 0.5 M NaCl. LTC_4 synthase was eluted with 300 mM imidazole, 0.5 M NaCl and 0.1 mM GSH in buffer A. The final step of purification was performed on a column packed with 3 ml S-hexylglutathione agarose. The column was washed with buffer A, supplemented with 0.5 M NaCl and 0.1 mM GSH. Pure LTC_4 synthase was eluted with 25 mM Tris-HCl, pH 7.8, 0.1% Triton X-100, 30 mM probenecid, 5 mM 2-mercaptoethanol and 0.1 mM GSH. The purified protein was either stored frozen at –20 °C or directly further polished in a buffer exchange step on a Superdex 200 16/60 (GE Healthcare) equilibrated with 0.03% w/v DDM (w/v), 20 mM Tris pH 8.0, 300 mM NaCl and 0.5 mM TCEP. Fractions containing LTC_4 synthase were concentrated to 3.1 mg ml^{-1} by ultrafiltration.

Crystallization. Crystals were grown at either 4 or 20 °C from a 3.1 mg ml^{-1} membrane protein solution, using a sitting drop vapour diffusion technique. Crystals typically appeared after 3–4 d, reaching optimal size after approximately 7 d. Protein solution containing 20 mM Tris pH 8.0, 300 mM NaCl and 0.03% (w/v) DDM was mixed (1:1) with reservoir solution containing either 200 mM NaCl, 100 mM Na cacodylate pH 6.5, 2 M $(NH_4)_2SO_4$ for native protein; 2% PEG 400, 100 mM HEPES-Na pH 7.5, 2 M $(NH_4)_2SO_4$ for GSH derivative or 100 mM bis-Tris pH 5.5, 2 M $(NH_4)_2SO_4$ for heavy atom soaks. For the latter, crystals were grown at 20 °C and soaked for 2 h in 2.5 mM $PtCN_4$, dissolved in artificial mother liquor. All crystals were transferred to their corresponding reservoir solution supplemented with 25% glycerol for cryo protection, then flash frozen in liquid nitrogen.

All X-ray data were collected on beam lines ID14-4 and ID23-2 at the European Synchrotron Radiation Facility (ESRF). Diffraction data of native and GSH soaked crystals was processed and scaled using Mosflm[17] and SCALA[18] while the XDS[19] suite was used for heavy atom MAD data sets. Three highly redundant data sets to 3.2 Å were collected around the Pt LIII edge. A single platinum site was located in the peak data set using SHELXD after local scaling in XPREP and SHELXE[20] had confirmed the correct hand. Experimental MAD phases were obtained using the program SHARP[21], resulting in an overall figure of merit (FOM) of 0.41/0.26 for the acentric/centric reflections respectively. These phases were further improved using a 75% solvent content in SOLOMON[22], resulting in an FOM of 0.88. The initial map was used to model α-helices, which were in turn used together with high-resolution data to assign residues with ARP/wARP[23]. The model was further built using Coot[24] and used for molecular replacement with Phaser[25] of a higher resolution apo structure and the GSH bound structure. Both apo and GSH bound structures were built in Coot and refined using Refmac[26] to an R_{work}/R_{free} of 17.6/21.2% and 18.3/22.0%, respectively. For further details, see Supplementary Table 1. All images were prepared using Pymol[27] except Supplementary Fig. 4, which was generated with LIGPLOT[28].

naturejobs

JOBS OF THE WEEK

During the 23 July US Democratic presidential debate, candidates responded to video questions from the public submitted via YouTube. In one video, an animated snowman begged the candidates to consider climate change, which it described as the "single most important issue to the snowmen of this country." Turning to a smaller snowman, it asked the candidates: "What will you do to ensure that my son will live a full and happy life?"

It was part of an innovative debate format that included several minutes of discussion about global warming, energy efficiency and nuclear power. Triggered in part by former presidential candidate Al Gore and his Oscar-winning movie *An Inconvenient Truth*, and by coverage of Intergovernmental Panel on Climate Change reports on global warming, these topics are receiving a spike in attention from the US public and policymakers.

What is the implication for research careers? With so much talk of innovative science and policy solutions for global warming, scientists should be looking at a better job market. But beware. As this week's feature reveals on page 618, it depends on where you look. The most promising jobs, in the short term, seem to be in the private sector, where companies are increasingly looking to hire climate experts. This makes sense — companies are under pressure to develop 'greener' practices that reduce carbon-dioxide emissions and foster sustainability. It is partly an attempt to be more responsible, and partly a public-relations ploy to appease customers and stockholders.

US academic institutions recognize the need for more training and the development of new interdisciplinary programmes, but many have not increased the number of tenure-track professorships in the atmospheric sciences. There are mixed prospects in Europe too, where, despite funding from the EU's Seventh Research Framework Programme, climate-science positions are surging only in niche areas. The YouTube generation has helped promote the issues, but young researchers interested in atmospheric science should be careful when considering their career paths.

Gene Russo, acting editor of *Naturejobs*

CONTACTS
Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

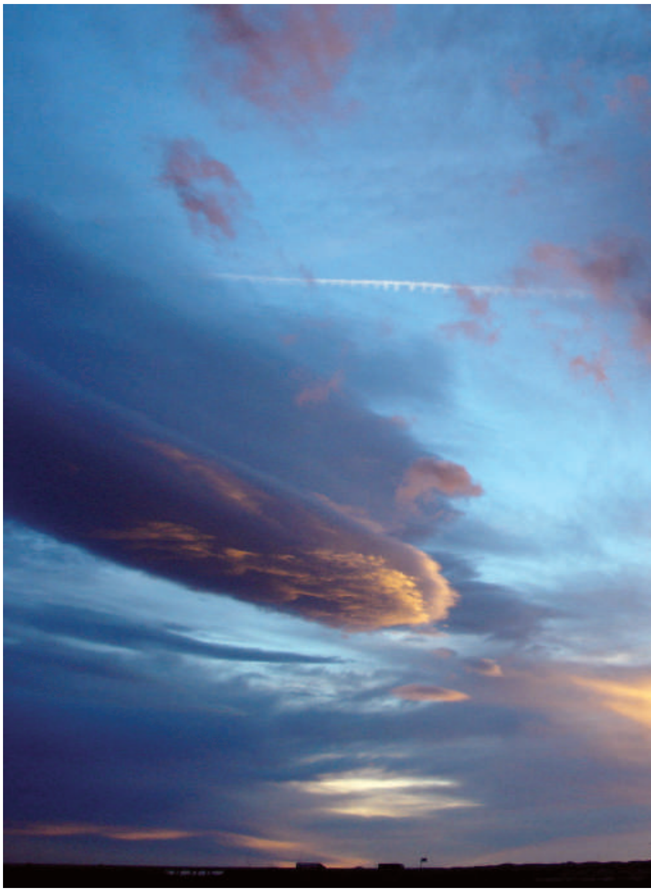
US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

CLIMATE OF OPPORTUNITY

Awareness about climate change is at an all-time high. Will this surge of attention translate into more jobs for climate scientists? **Amanda Haag** reports.



With global warming gracing covers from *Rolling Stone* to *Sports Illustrated*, it is clear that extreme weather, climate change and environmental concerns have penetrated the mainstream psyche in the United States. Terms such as 'green', 'sustainable', 'energy efficient' and 'carbon neutral' are now in vogue. Although these have been active research areas for years, heightened awareness is already affecting jobs and career prospects.

Graduates and faculty members alike are testing the waters in new interdisciplinary arenas and the private sector and local governments are looking for climate-science researchers to apply their skills in new ways.

In industry, various new job descriptions have been created. Large corporations are now adding chief sustainability officers to their line-ups. Google boasts a 'green energy czar', and Home Depot — a home improvement retailer — now has a vice-president for environmental innovation. As no comprehensive training exists for such roles, hybrid degree programmes are being founded to produce a workforce that's qualified to fill them. "Global warming and the need to become a sustainable society are driving a new industrial economic revolution," says Tom Kimmerer, executive director of the Association for the Advancement of Sustainability in Higher Education, based in Lexington, Kentucky. "Any time you have an industrial economic revolution, you create thousands of new jobs in disciplines that didn't exist before, and I think we're at the early stages of that."

There is great demand for graduates who have grappled with the complexity and uncertainty of issues such as environmental sustainability, development, economics and climate change, says Charles Redman, director of the new School of Sustainability at Arizona State University in Tempe, which will welcome its first class this autumn (see 'Lessons in climate').

UNIV. CALIFORNIA, BERKELEY

Ronald Cohen is positive about the opportinities for climate-science graduates

"We've talked to companies such as Wal-Mart, Starbucks, British Petroleum and U-Haul, each of whom already has an officer or two to spearhead this work," Redman says. "After the visit they say, 'do you have people ready for us to hire? We need people right now.'" Redman points out that even though the United States has not ratified the Kyoto Protocol, the mayors of more than 600 US cities have signed on to meet its requirements, and many are now looking for environmental leaders at the state and local levels to set them on the right path. And Ronald Cohen, director of the Berkeley Atmospheric Science Center at the University of California, says that graduates of his programme are being sought to fill positions at the intersection of policy and science at places such as the California Air Resources Board, which will implement the state's strict laws for limiting emissions from vehicle exhausts.

LESSONS IN CLIMATE

A handful of new programmes are springing up to meet the demand from industy for experts in everything from product sustainability to carbon trading and carbon neutrality.

At the University of Edinburgh, UK, faculty members from the geociences school have teamed up with the business school to design an MSc in carbon management, a one-year masters programme that will kick off in the autumn of 2008. Students will take core courses in carbon trading, the carbon cycle, carbon legislation, and climate change impacts and adaptation. They will have opportunities to specialize in areas from risk analysis to economics of the energy industry.

"The demand [for graduates of these courses] is there and we need to fill it quite urgently," says David Reay, a climate scientist who is director of the programme.

Likewise, the University of East Anglia, UK, has just launched the world's first MBA in carbon management, a one-year programme that will begin accepting students in January 2008. **A.H.**

More undergraduates than jobs

In the world of traditional academia, the job market for positions that criss-cross the environmental and atmospheric sciences isn't clearly defined. According to an as-yet unpublished study of undergraduate enrolment and job trends in meteorology, five to ten times more US undergraduates are enrolling in meteorology courses than there are jobs for them. "If we don't have a problem now, we could have a problem later on as students who can't get jobs at the bachelor's level go on to graduate school," says study author John Knox, a meteorologist at the University of Georgia, Athens. (see 'Mixed prospects').

Positions open to bachelor's recipients include broadcast meteorology for local weather stations as well as forecasting and consulting for the more than

C. STONEBURNER

NASA

NOAA/OAR/ERL/NSSL

Broadcast meteorology is one of the many career opportunities available.

250 private US meteorological companies.

And the atmospheric science community in the United States is concerned about flagging funding from key government agencies such as the National Oceanic and Atmospheric Administration (NOAA), NASA and the Department of Defense. In fiscal year 2004, 51% of the funding for weather and space-weather research came from NASA, 20% from NOAA, and 12% from the defence department, according to a 2007 report from the US National Academy of Sciences. But over the past five to ten years, NASA and the defence department have significantly decreased their support for basic atmospheric-science research, and NOAA has faced budget constraints in various Earth and atmospheric science departments for years. Richard Anthes, president of the University Corporation for Atmospheric Research, based in Boulder, Colorado, estimates that NASA's budget has dropped in terms of real purchasing power by some 35% in the past three to four years. The end result, according to the National Academy of Sciences' report, is that the research community will increasingly look to the National Science Foundation for funding, making research grants even harder to come by.

But there are opportunities. NOAA's National Centers for Environmental Prediction hires about 40 new people annually, or some 10% of its workforce, according to the centre's director, Louis Uccellini. Bachelor's or master's-level degree recipients generally hold forecast desk positions, and doctorates fill 95% of research positions. Uccellini stresses that in every arena, from basic forecasting to research monitoring and data assimilation, the desired candidates are those with an eye towards an interdisciplinary, 'Earth-systems' approach. "Coupling the interactions of ocean, ice and land to atmosphere is a huge issue for weather and climate prediction," says Uccellini. "We're just on the front edge of that."

In academia, government, the private sector and myriad non-profit organizations, the boundaries between climate change, atmospheric science and environmental sustainability continue to blur. And with that comes no shortage of creative career moves within and between fields. "You're going to see a lot of opportunities for people to retool their careers," Kimmerer says. ■

Amanda Haag is a freelance science writer based in Denver, Colorado.

MIXED PROSPECTS

Climate-change science has long been a part of the mainstream in Europe. The European Union's Seventh Research Framework Programme includes €1,890 million (US$2,590 million) for environment, sustainability and climate change research for 2007–13 — a 40% rise above previous funding.

And yet job prospects are mixed. For example, in Britain, as in the United States, demand is growing within the private sector for climate expertise and for scientists able to bridge disciplinary boundaries. The Met Office, the country's national weather service, hears concerns about climate-change effects from representatives within the construction, insurance and energy industries, says Wayne Elliott, business manager at the Met Office. "With severe weather impacting on the bottom line to a large degree, we're being asked to answer questions that we haven't even addressed before," says Elliott. Companies need to know how climate will affect them, and how they should prepare and adapt. He notes that one of the Met Office's leading climate scientists recently moved into the service's business section to work directly with the insurance and energy industries.

A. ZIMDARS

Paul Williams is concerned that there is a lack of university jobs in Britain.

Some doctorates and postdoctoral fellows are already sensing a scarcity of permanent faculty positions. Paul Williams, a postdoctoral research fellow in atmospheric modelling at the University of Reading, is worried about a shortage of university jobs in atmospheric science, a relatively young discipline offering fewer positions than traditional departments such as maths, physics and chemistry.

Graduates of these traditional departments are also moving into the atmospheric–sciences arena to apply their training, further limiting opportunities.

Neither is Germany is experiencing a surge in climate-science roles, says Jochem Marotzke, a director of the Max Planck Institute for Meteorology in Hamburg, because interest in basic climate science peaked several years ago. Yet a clear need is developing for detailed climate information for local governments. He says there's "a lot more work to be done before we can provide this information reliably". Germany is focusing largely on mitigation and adaptation issues and analysing what directions this work should take, he says. To that end, German research minister Annette Schavan announced in February that €255 million would be allocated to a climate 'action programme' over the next three years, including expansion of the German Climate Computing Centre, Hamburg. How much money will be spent on classical research versus adaptation efforts — and how these will translate into new job opportunities — is still under discussion. **A.H.**

National Center for Complementary and Alternative Medicine
HEALTH SCIENTIST ADMINISTRATORS

The National Center for Complementary and Alternative Medicine (NCCAM), a component of the National Institutes of Health (NIH) in the Department of Health and Human Services (DHHS), is seeking applications from exceptional scientists to serve as Health Scientist Administrators (HSAs) in its Division of Extramural Research and Training. As an HSA, you will develop a strategic vision for complementary and alternative medicine research; implement plans to achieve that vision; provide stewardship of awarded grants and contracts; and serve as a scientific expert within your field.

Qualifications: The successful individuals will possess a Ph.D., M.D., or equivalent degree and a record of research accomplishment in the areas of neuroscience, molecular biology, biochemistry, chemistry, pharmacology, and/or clinical research; the successful candidates are also expected to demonstrate strategic vision; excellent communication skills; and the ability to work collaboratively in a team.

Salary: The current salary range is $79,397.00 - $121,967.00, depending on experience and accomplishments; a full Civil Service package of benefits (including retirement, health, life and long-term care insurance, Thrift Savings Plan participation, etc.) is available. Recruitment incentive may be awarded and relocation expenses may be paid.

How To Apply: Position requirements and detailed application procedures are provided in vacancy announcements **NCCAM -07-193989-CR-MP and NCCAM-07-193989-CR-DE**, which can be found at **http://nccam.nih.gov**. All applications and supplemental information must be received no later than **Friday, August 3, 2007**. For additional information, contact **Ms. Kathy Delauter at 301-594-2283 or delauterk@mail.nih.gov.**

TENURE TRACK POSITION
LABORATORY OF MOLECULAR BIOLOGY

The Laboratory of Molecular Biology (LMB), Center for Cancer Research, of the National Cancer Institute, National Institutes of Health (http://ccr.cancer.gov/labs/lab.asp?labid=99) uses genetics, molecular biology, cell biology, and molecular modeling to examine and solve a broad range of important biological problems. The Laboratory now invites applications for a tenure track position in the Laboratory of Molecular Biology, CCR, NCI for a scientist working in the field of antibody engineering as it relates to cancer therapy. Candidates must have a Ph.D. or M.D. and a proven record of innovative research and productivity in antibody engineering, display of antibody fragments on phage and mammalian cells and cancer therapeutics. The successful candidate will join an active group of translational and clinical investigators in the LMB working on immunotoxins, humanized antibodies, and toxin biology and carrying out clinical trials with immunotoxins and antibodies. Salary will be commensurate with education and experience. A two-page statement of research interests and goals should be submitted in addition to three letters of recommendation and a curriculum vitae to **Mrs. Ann Schombert, Executive Secretary, Laboratory of Molecular Biology, CCR, NCI, Building 37, Room 5106, Bethesda, MD 20892-4264; phone: 301-451-8714, Fax: 301-402-1344, email: schombea@pop.nci.nih.gov**. NIH Tenure track investigators with educational debts may be eligible for the NIH Loan Repayment Program. The NCI is an Equal Opportunity Employer. The closing date for applications to be accepted is **September 15, 2007**.

Biostatistician

The intramural program of NINDS is seeking an experienced Biostatistician to serve in its Biostatistics Unit at the NIH Clinical Center in Bethesda, Maryland beginning in August 2007. This is a full-time position with responsibility for reviewing and providing statistical support toward improving clinical protocols; general statistical consultation service to research staff for data analysis and study design; and substantial collaboration on projects as a coauthor in published work.

A Doctoral Degree is required. Experience in biostatistics or statistics with working experience involving biostatistical applications to clinical trials and/or observational studies is preferred. Appointment will be in the Staff Scientist (non-tenure track) series and salary will be commensurate with experience.

Applicants should send a cover letter, C.V., and names of three references to:

Ms. Caren Collins, NINDS/NIH/DHHS, 10/5N254, 10 Center Dr., MSC 1430, Bethesda, MD 20892, or via E-mail: collinsca@ninds.nih.gov.

The National Institute of Neurological Disorders and Stroke (NINDS) is a component of the National Institutes of Health (NIH) and the Department of Health and Human Services (DHHS). DHHS and NIH are Equal Opportunity Employers.

NW108353R

THE NIH IS DEDICATED TO BUILDING A DIVERSE COMMUNITY IN ITS TRAINING AND EMPLOYMENT PROGRAMS

OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH

Junk

Be careful what you throw away.

Gord Sellar

SIR — After reading your publication's special issue focused on the debate over the 'junk-DNA babies' and GM research, I felt obligated to comment.

I am a genetics-therapy counsellor. People come to me everyday, asking: "Why can't I have kids? Why me?" The less fortunate ask: "Why did my kid end up like this? Don't you people know what you're doing?"

The suffering these latter parents undergo is unimaginable. The genetic and developmental illnesses of the past — for example, thalidomide babies, autism and fetal alcohol syndrome — seem almost trivial compared with the conditions of babies born to unknowing parents who underwent mainstream genetic therapy during the early biotech boom a few decades ago.

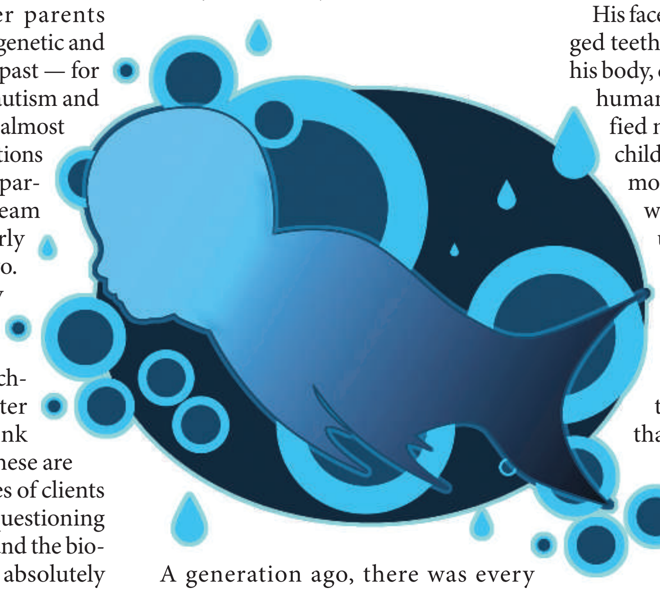
Imagine your newborn baby covered with thick fur or a chitinous carapace or extra, non-human limbs. Imagine catching your two-year-old daughter laying small, thumb-sized, pink eggs on her playroom floor. These are just a few examples of real cases of clients I've counselled myself. Their questioning of the scientific establishment and the biotech industry's competence is absolutely understandable.

I tell them that we do know what we're doing *now*, much more than we did a generation ago. That we have learned from the mistakes we've made, and will not repeat them.

This answer hasn't satisfied anyone yet, but it's the truth. For all their children's suffering, the parents who come to see me have led very normal, healthy lives. That they remained unaware of their genetic treatments before pregnancy demonstrates how successfully their modifications warded off disease and minimized contact with the medical establishment. Is involuntary childlessness too high a price for them to pay, given such benefits?

A man who had undergone pre-emptive genetic alteration visited me the other day. His alterations were aimed at preventing early-onset Alzheimer's, a profoundly disruptive condition that would have directly affected the quality of his life. It took some time, but finally he agreed that the alterations made to his genes had been effective, and truly had benefited him. Unfortunately, they had been made before the differentiation of his sex cells.

But at the time when he was conceived, the only treatment method technologically and economically available was to modify his genes before cell differentiation. The stripping of 'junk DNA' and its replacement with coded marker sequences was the easiest way to ensure quick, effective deployment of the viral reverse-transcriptase payload and the easiest way to target only the genes needed. If only 'junk' DNA had actually been mere junk.



JACEY

A generation ago, there was every indication that it was. Noncoding DNA had (and still generally has) zero apparent impact on the functioning of somatic cells. Plenty of species, such as copepods, certain roundworm species and sciarid flies, have been naturally excising their noncoding DNA when forming their somatic cells for hundreds of millions of years without problems of this kind, for example. And in the short term, it has been an excellent method, which is why, even as scientists began to speculate about the role of junk DNA in long-term genomic stability, the industry of genetic treatment continued as it had done for years. Short-term success simply wouldn't heed the alarms of long-term problems, until it was too late.

I understand their plight, and the stigma they suffer under, better than most people can. I face it everyday myself. I was treated *in vitro* for several familial cancer predispositions, and for other specific long-term epigenetic factors. Like many, I remained unaware of this fact until an ultrasound revealed my son's condition to my wife and me.

My son is now three years old. He was born armless, legless, with fins and a set of gills on his neck, and after he was born, he began to drown in the dry air of the delivery room. My ex-wife, whose genes were unmodified, was not shown him until hours after his birth. We had been warned of his condition months before, but there are things for which it is impossible adequately to brace oneself. I doubt that either of us shall ever fully recover from the shock.

His face, with its wide black eyes and jagged teeth, the mottled blue-black skin of his body, combined with the unmistakably human sound of his infant cry horrified me, and I instinctively hated my child. The memory of that feeling, in a moment that should have been filled with joy, is a burden that will weigh upon me for the rest of my life.

But I have adjusted, and now I care for my son in his mother's absence — like many, she was understandably too distraught to parent a child such as Christopher. I have a tank in my home that is large enough to accommodate him, and through our awkward mode of communication, I try to show him love. I see intelligence in his eyes, a very human intelligence, but I have, finally, given up all hope of communication.

Yet I do not blame my parents, or the scientists of the past. It's wrong to criticize people who long ago had their children's genes modified. After all, they were all simply doing the best they could for us, their children, at the time.

This all shows that more, not less, research is needed. Caution is indeed important, but we must not cut off research funding. Those who call for a moratorium are mistaken in thinking that tragedies will be averted; turning our backs on research will result in more such tragedies. Despite past mistakes, we must not abandon our quest to understand our own genome, who we are, and what we are to become.

Sincerely,
Trevor Chilsom,
Federal Genetics Treatment Counseling & Research Center, Rochester, New York. ■

Gord Sellar is a Canadian who has been living and working in South Korea since 2002. He is a graduate of Clarion West 2006.

FUTURES